Using Computers to Develop Phonemic Awareness in the Early Primary Classroom

Linda Snape

Doctor of Philosophy Submission: January 2011

Using Computers to Develop Phonemic Awareness in the Early Primary Classroom

BY

LINDA SNAPE

UNIVERSITY OF CENTRAL LANCASHIRE

Thesis submitted to the University of Central Lancashire in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

January 2011

The work presented in this thesis was carried out in the School of Computing, Engineering and Physical Sciences at the University of Central Lancashire.

Declaration

I declare that while registered with the University of Central Lancashire for the degree of Doctor of Philosophy I have not been a registered candidate or enrolled student for another award of the University of Central Lancashire, or any other academic or professional institution during the research programme. No portion of the work referred to in this thesis has been submitted in support of any application for another degree or qualification of any other university or institution of learning.

Linda Snape

Abstract

The aim of this project is to determine whether a computer application can be used to develop phonemic awareness in the early primary classroom, which is a key component of phonics. This thesis explores the evolution of the strategy for teaching literacy in the UK which shows phonics to be a key component of that strategy. However, government reports which inform the direction of the literacy strategy call for more empirical study in all areas of literacy teaching; this thesis documents such an empirical study.

This research project creates a phonics-based computer application designed specifically for young children aged 5 to 6 years (year 1 in UK primary schools). The timing and level of content presented by the computer application activities are grounded in appropriate academic theory. A significant component of the work is the development of interface design guidelines for children's applications. These guidelines are then used to inform the development of the phonics-based computer application. A Randomised Controlled Trial (RCT) is designed to determine the application's effectiveness in developing the phonemic awareness skills of young children in a classroom setting. In order to control experimental bias resulting from problems with the usability of the computer interface, the usability of the application's interface is evaluated in the classroom by year 1 children before the application is used in a pragmatic RCT. The results of the final usability evaluation found no usability issues and the application was wholly intuitive to the children in the evaluation groups.

The results from the RCT (N=266) show no statistically significant improvement in the learning rate of phonemic awareness by the intervention group using the computer program compared to the traditional teacher-delivered paper-based method used with the control group, even though the computer program was designed carefully for this age range. The results did suggest however, that the intervention group developed at the same rate as the control group which implies that the computer program could be used to support teachers by reducing the amount of resource-intensive phonics tuition required by children in this age range.

Acknowledgements

I would like to thank my supervisory team

Dr Martin Brown

Dr Stuart MacFarlane

Dr Chris Smith

and

Mr John Paine

for their valuable time and guidance.

I am indebted to my colleagues

Dr Antony Nicol

and

Mr Chris Casey

for their invaluable feedback, support and encouragement.

I would particularly like to thank my children, Katie and Ryan for their patience during the period of this project.

Contents

CHAPTER 1 INTRODUCTION AND THESIS STRUCTURE....1-2

1.1 High Level Thesis Structure 1-2
1.2 Introduction.... 1-4
1.2.1 Evaluating the usability of user-centric software applications.... 1-4
1.2.2 Evaluating the educational effectiveness of user-centric applications.... 1-5
1.3 Detailed Thesis Structure 1-6

CHAPTER 2 COMPUTERS IN THE CLASSROOM.... 2-2

2.1 Introduction.... 2-2
2.2 Can computers be used to support learning in the primary classroom?.... 2-2
2.3 Is ICT useful and effective for reading and spelling in the classroom? 2-3
2.4 What do results from evaluations of phonics computer programs indicate?.... 2-6
2.5 Are teachers a barrier to the use of computers in the early primary classroom?.... 2-8
2.6 Is gender a barrier to the use of computers in the early primary classroom?.... 2-9
2.7 Is age a barrier to computer-based learning in the early years? 2-10
2.8 Conclusions 2-11

CHAPTER 3 PHONICS IN UK CLASSROOMS: THE DEBATE.... 3-2

3.1 Introduction.... 3-2
3.2 Phonics – an overview 3-2
3.2.1 Synthetic phonics 3-3
3.2.2 Analytic phonics 3-3
3.3 The development of Phonics teaching.... 3-4
3.3.1 1989 - 1998 3-6
3.3.2 1998 3-7
3.3.3 1999 3-8
3.3.4 2002 3-9
3.3.5 2003 3-10
3.3.6 2004 3-12
3.3.7 2005 3-12
3.3.8 2006 3-13
3.3.9 2007 3-16
3.4 Conclusions 3-17

CHAPTER 4 RESEARCH AND EVALUATION METHODS 4-2

4.1 Introduction.... 4-2
4.2 Quantitative research method.... 4-2
4.2.1 Chance bias.... 4-3
4.2.2 Gender bias 4-3
4.2.3 Age bias.... 4-4
4.2.4 Selection / Familiarity bias 4-4
4.2.5 Contamination bias.... 4-5
4.2.6 Performance bias.... 4-5
4.2.7 Dilution bias.... 4-5
4.2.8 Demoralisation effect.... 4-6
4.2.9 Reporting bias 4-6
4.2.10 Attrition bias.... 4-6
4.2.11 Bias control in the SPEL RCT 4-8
4.3 Qualitative evaluation methods 4-8
4.3.1 Observation 4-10
4.3.2 Action tracking 4-10
4.3.3 Thinking aloud, Constructive Interaction and Co-discovery.... 4-10
4.3.4 Performance measures.... 4-12

4.3.5 Retrospection and post-session interviews ... 4-12
4.3.6 Post-Evaluation Questionnaires and Interviews ... 4-13
4.4 RESEARCH ETHICAL CONSIDERATIONS ... 4-13
4.5 CONCLUSIONS ... 4-14

CHAPTER 5 SPEL OPERATION AND INTERFACE DESIGN ... 5-2

5.1 INTRODUCTION ... 5-2
5.2 GENERATING NATURAL HUMAN SPEECH OUTPUT ... 5-2
5.3 IMPLEMENTATION OF CHILD COMPUTER INTERACTION GUIDELINES ... 5-4
5.4 SPEL ARCHITECTURAL OVERVIEW ... 5-7
5.4.1 SPEL High level technical details ... 5-7
5.4.2 SPEL Sequencing and Segmenting Activities ... 5-9

CHAPTER 6 SPEL USABILITY AND PHONEMIC AWARENESS EXPERIMENTS ... 6-2

6.1 INTRODUCTION ... 6-2
6.2 SPEL USABILITY EXPERIMENT ... 6-3
6.2.1 SPEL usability issues from iteration 1 ... 6-6
6.2.2 SPEL usability issues from iteration 2 ... 6-9
6.2.3 Evaluation feedback interviews ... 6-9
6.2.4 Observational findings ... 6-11
6.2.5 Usability Conclusions ... 6-12
6.3 PRAGMATIC RANDOMISED CONTROLLED EXPERIMENT ... 6-14
6.3.1 Pilot study summary ... 6-14
6.3.2 A large scale RCT ... 6-17
6.3.2.1 Power calculation for the SPEL randomised controlled trial ... 6-17
6.3.2.2 Experiment design detail ... 6-21
6.3.2.3 Method of randomisation ... 6-23
6.3.2.4 Experimental process ... 6-25
6.3.2.5 Tests ... 6-26
6.3.2.6 Intervention and control treatment ... 6-29
6.3.2.7 Word Lists ... 6-30
6.3.3 SPEL Pragmatic Randomised Controlled Trial results ... 6-30
6.3.3.1 Attrition ... 6-32
6.3.3.2 Summary of results ... 6-33
6.4 CONCLUSIONS ... 6-35

CHAPTER 7 CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER WORK ... 7-2

7.1 THESIS AIMS ... 7-2
7.2 CONCLUSIONS ... 7-2
7.2.1 Problems and benefits of using computers as a teaching aid in the early primary classroom ... 7-3
7.2.2 To understand the approach to teaching phonics in the early primary classroom ... 7-3
7.2.3 Produce and evaluate an intuitive user interface for young children ... 7-4
7.2.4 Implement an effective educational tool for primary classroom teaching ... 7-5
7.2.5 Carry out an educational effectiveness RCT ... 7-6
7.3 THESIS CONCLUSION ... 7-6
7.4 ORIGINAL CONTRIBUTIONS TO KNOWLEDGE FOR THE RESEARCH COMMUNITY ... 7-7
7.5 RECOMMENDATIONS FOR FURTHER WORK ... 7-8

REFERENCES ... R-2

APPENDIX A CHILD-COMPUTER INTERACTION ... A-2

A1 INTRODUCTION ... A-2
A2 DESIGN GUIDELINES DEVELOPMENT ... A-3
A2.1 Minimise the use of text as a feedback mechanism ... A-7
A2.2 Minimise the number of interactive controls ... A-8
A2.3 Vary the feedback ... A-10
A2.4 Minimise keyboard input ... A-10
A2.5 Make use of randomness ... A-12
A2.6 Enable early exit from the application which is inaccessible to children ... A-12
A2.7 Provide tiered pro-active interactive help ... A-12
A2.8 Enable user defined options but make them inaccessible to the children. ... A-14

A2.9 Ensure that interaction hardware is of an appropriate size for the user *A-15*
A2.10 Design for no scrolling *A-15*
A3 CONCLUSIONS A-15

APPENDIX B LIST OF ENGLISH GRAPHEMES AND PHONEMES B-2

APPENDIX C USABILITY EVALUATION QUESTIONNAIRE C-2

APPENDIX D TEST SCORE SHEET AND INSTRUCTIONS D-2

APPENDIX E SPEL APPLICATION CONFIGURATION E-2

APPENDIX F RANDOM BOOLEAN ALGORITHM F-2

APPENDIX G CAPTURING INTERACTION AND PERFORMANCE DATA G-2

G1 OVERVIEW G-2
G1.1 Capture *G-3*
G1.2 Id Number *G-3*
G1.3 Session Number *G-3*
G1.4 Object *G-3*
G1.5 Downtime *G-3*
G1.6 DownX and DownY *G-3*
G1.7 DownString *G-4*
G1.8 UpTime *G-4*
G1.9 UpX and UpY *G-4*
G1.10 UpString *G-4*
G1.11 Button *G-4*

APPENDIX H STATISTICAL ANALYSIS DETAILS H-2

H1 OVERVIEW H-2
H2 CHECK FOR NORMALITY OF THE SAMPLE H-2
H3 TEST MEANS AND STANDARD DEVIATIONS H-3
H4 ANCOVA H-4
H5 INDEPENDENT SAMPLES T-TEST H-4
H6 EFFECT SIZE CALCULATIONS H-5
H7 ATTRITION TEST H-6
H8 SPSS PROGRAM CODE TO REPLICATE TESTS H-8

APPENDIX I RESULTS SPREADSHEET EXAMPLES I-2

Figures

Figure 1-1: Thesis Structure 1-3
Figure 3-1: Phonics' development timeline 3-5
Figure 3-2: The NLS Searchlights Model 3-8
Figure 3-3: Simplified reading model 3-15
Figure 4-1: "Smileometer" questionnaire gauge 4-13
Figure 5-1: SPEL application block diagram 5-7
Figure 5-2: Example database table created by the Sequencing activity 5-8
Figure 5-3: Conceptual schema of synthetic phonics 5-9
Figure 5-4: Sequencing the word "like" 5-11
Figure 5-5: Segmenting activity breaking down the word "saw" 5-12
Figure 6-1: Interface usability evaluation lifecycle 6-4
Figure 6-2: Child waiting for instructions to finish 6-8
Figure 6-3: SPEL questionnaire results 6-11
Figure 6-4: Power calculation results from G*Power 6-18

Figure A-1: Letterworld home screen and one of the letter formation activities A-5
Figure A-2: Literacyworld home screen and one of the activity screens A-5
Figure A-3: Vocabulary Tutor activity screen A-5
Figure A-4: SPEL session during evaluation A-6
Figure A-5: Alphabet Activity Screen A-9
Figure A-6: User defined options interface used in the Letterworld prototype A-14
Figure D-1: Pupil Score Sheet D-2
Figure E-1: Activity / word trigger dates E-2
Figure E-2: Activity sequence based on date triggers E-3
Figure E-3: Words0 list E-4
Figure E-4: Words5 list E-4
Figure E-5: Partial listing of the \Word directory E-5
Figure E-6: Words0 digraphs E-6
Figure E-7: Words5 digraphs E-6
Figure E-8: Teacher activity override screen E-7
Figure E-9: SPEL directory structure E-8
Figure G-1: Example data capture table G-2
Figure G-2: Data capture from a stronger pupil G-5
Figure H-1: Histogram of all participants' pre-test scores H-2
Figure H-2: Q-Q plot for control group pre-test scores H-3
Figure H-3: Means after extreme adjustment for attrition H-7
Figure H-4: ANCOVA results after extreme adjustment for attrition H-7

Tables

Table 5-1: Application speech script example 5-3
Table 6-1: Phoneme-to-Grapheme descriptive statistics 6-15
Table 6-2: Grapheme-to-Phoneme descriptive statistics 6-15
Table 6-3: Mean and standard deviation of each group 6-31
Table 6-4: Attrition profile 6-32
Table 6-5: Summary of RCT results 6-35

Table H-1: Pre-and Post-test means and standard deviations H-3
Table H-2: ANCOVA on post-test scores using pre-test scores as a covariate H-4
Table H-3: Pre- and post-test means and standard deviations H-5
Table H-4: T-test comparing Pre- and Post-test scores H-5

Chapter 1

Introduction and Thesis Structure

Chapter 1 Introduction and Thesis Structure

1.1 High Level Thesis Structure

This thesis aims to determine whether a computer application can be used to develop phonemic awareness in the early primary classroom. An application, the System for Phonics Early Learning (SPEL), was developed and evaluated to facilitate this aim.

A literature review was undertaken to determine whether others had carried out a similar study and to determine whether computers are considered to be a suitable learning vehicle for the early primary classroom; this is covered in detail in **Chapter** 2 - ***Computers in the Classroom***. The development of a phonics-based application needs underpinning by relevant theory which is discussed in **Chapter 3** - ***Phonics in UK classrooms: the debate***. A user-centric application is only as good as its usability. To ensure the application was suitable for children as young as 5 years, it was necessary to research and develop the area of Child-Computer Interaction (ChiCI) as this field was in its infancy. **Chapter 5** - ***SPEL Operation and Interface Design*** which is detailed in **Appendix A** - ***Child-Computer Interaction*** discusses the work undertaken and contributions to knowledge in this relatively new field; contributions include peer reviewed research papers and a set of interaction design guidelines. With the SPEL application built on academic principles and newly developed interface guidelines, the research approach was designed. **Chapter** 4 - ***Research and Evaluation Methods*** details the approach used to evaluate SPEL's usability and determine its educational effectiveness in the classroom. The implementation of three experiments: a qualitative usability evaluation of the SPEL interface; a pilot Randomised Controlled Trial (RCT) and a large scale RCT are discussed and results detailed in **Chapter 6** - ***SPEL Usability and Phonemic Awareness Experiments*** and **Appendix H** - ***Statistical Analysis Details***. A key contribution to knowledge in this section is the detailed experimental process, a substantial data collection and results of one of the largest experiments of its kind. **Chapter 7** - ***Conclusions and Recommendations for Further Work*** reflects on what has been achieved and how this work and use of the data from a battery of literacy-based tests can be of further benefit to the research community. Figure 1-1 illustrates the thesis structure and indicates key appendices containing details of related chapters; as indicated, some areas can be bypassed without losing the flow of work.

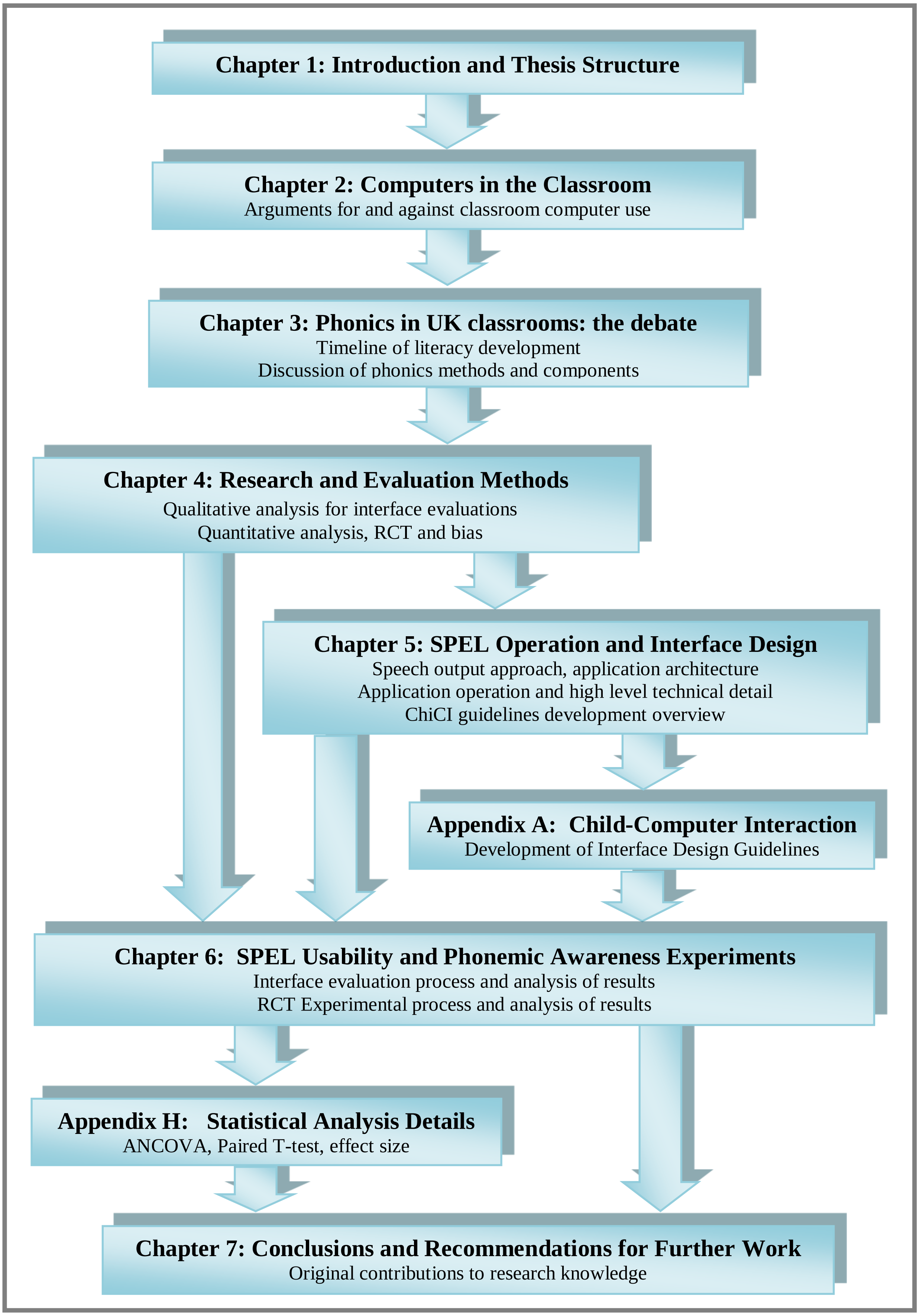

Figure 1-1: Thesis Structure

1.2 Introduction

This thesis details the process of developing and evaluating the usability and educational effectiveness of a custom built phonics-based computer application in the early primary classroom. Two distinct research approaches were required to achieve this: qualitative research techniques were used to determine the usability of the SPEL application which involved the study and development of the Human-Computer Interaction (HCI) field of research and a quantitative approach based in the area of scientific experimentation using an RCT was used to determine the educational effectiveness of SPEL. These two distinct approaches are introduced in this chapter then detailed in subsequent chapters.

To simplify reading this document, some conventions and terms have been adopted:

- **This project** or **this research** refers to the work undertaken and reported in this document.
- ↔ means bidirectional relationship. Rather than repeatedly writing sound to symbol mapping and symbol to sound mapping for example, this is abbreviated to sound↔symbol.
- Intention To Treat and Intention To Teach analysis refer to the same analysis but either Treat or Teach would be used in the relevant context; this document will simply abbreviate to ITT in all cases.

1.2.1 **Evaluating the usability of user-centric software applications**

HCI is concerned with the design, evaluation and interaction of computer systems. Usability is a term used to describe the success of a computer interface in terms of how well the users are able to complete their tasks and is concerned with factors such as the types of users, the types of task and hardware constraints (Leventhal and Barnes, 2007). Building usable interfaces is not easy and when the users are young children who are not yet able to read, a new set of challenges need to be addressed; an overview of the work carried out to develop the field of HCI for pre-readers is discussed in Chapter 5 - *SPEL Operation and Interface Design* and details for application developers are available in Appendix A - *Child-Computer Interaction*.

The development process and the elements that make up the final product have an influence on usability (Leventhal and Barnes, 2007). If a phonics computer application has not been evaluated from a usability viewpoint, the user interface cannot be discounted as a confounding factor in a scientific evaluation of the effectiveness of the application as a phonemic awareness teaching tool. It was therefore necessary to search for usability evaluations of existing phonics applications as background work to this project in an attempt to find an application which could be evaluated from this perspective. As no relevant usability study could be found, it was necessary to develop and evaluate a bespoke application as part of the project.

A phonemic awareness software application was developed and a formal evaluation of its high level of usability within a UK classroom setting was assured before using it in an RCT.

Details of the research approach used to evaluate the usability can be found in Section 4.3 - *Qualitative evaluation methods*. A report of the evaluations and results are recorded in Section 6.2 - *SPEL Usability experiment*. An overview of the SPEL application is discussed in Chapter 5 - *SPEL Operation and Interface Design* and the interface guideline development is detailed in Appendix A - *Child-Computer Interaction.*

1.2.2 **Evaluating the educational effectiveness of user-centric applications**

An RCT involving 266 children from four schools in the North West of England was carried out over a three month period to determine the educational effectiveness of the application as a phonemic awareness tutor; the application is known as the System for Phonic Early Learning - SPEL (Snape et al., 2003, Snape, 2007).

Brooks et al (2006) reported the largest UK-based RCT of Information and Communication Technologies (ICT) for the teaching of spelling and reading among students of school age. The RCT carried out as part of this PhD project will contribute to knowledge in this area by extending Brooks' work into the younger age group of 5 to 6 years. The results of this work will be useful to government bodies, educationalists and researchers involved in making recommendations for future educational policy regarding the introduction of computer software for phonemic awareness into mainstream schooling.

A further contribution to knowledge is a potential reduction in the cost of supervision for computer applications which have an interface which is intuitive to non- or early-readers. Because the intervention is to be delivered in the school computer suites with minimal staff supervision, it will provide a true reflection of the resource cost of delivering SPEL as computer-based support.

A key contribution to knowledge for the research community is the detailed recording of the RCT process which will enable replication of the results reported in this thesis and inform others of the meticulous process required to carry such a study to ensure validity of the results.

1.3 Detailed Thesis Structure

Chapter 1 sets the context and outlines the research goals. Chapter 2 - *Computers in the Classroom* outlines the views of researchers, teachers and the government to the introduction of ICT into the classroom and provides a review of the high-quality evidence available to support the use of ICT in primary schools in the UK. It comprises a literature review of research work carried out to determine whether ICT should be used as a matter of course in modern teaching with particular emphasis on young children. Using computer applications can be resource intensive for teachers as children often need help with many commercial applications as they can get lost in the navigation or simply don't know what to do next. The conclusions to Chapter 2 - *Computers in the Classroom* indicate the need for more empirical research into the effectiveness of computers as teaching tools with particular emphasis on a rigorous, replicable study design which can be used to inform future education policy in the UK. HCI is a key component in software development and the usability of computer interfaces should also be evaluated before trials are conducted to evaluate the effectiveness of computer applications in the classroom.

The manipulation of sub-word elements is identified as a key initial reading component. This is the subject of Chapter 3 - *Phonics in UK classrooms: the debate*, where phonics is introduced. The Phonics chapter outlines and explains the key components of phonemic awareness. The evolution of literacy teaching over the past twenty years is traced and summarised diagrammatically as a quick reference in Figure 3-1 on page 3-5. Phonics is a clear thread through the evolution and matures into the specific phonics approach used today in UK classrooms. The chapter discusses key government recommendations towards the teaching of phonics. These recommendations are used to inform the design of the SPEL application.

Chapter 4 - *Research and Evaluation Methods*, discusses a quantitative approach to evaluating the educational effectiveness of SPEL in developing phonemic awareness and a qualitative approach to determine the effectiveness of SPEL's interface such that its operation is so intuitive that children should rarely need to ask a supervisor for help. The chapter discusses the rigorous approach of a randomised controlled trial. A detailed discussion of potential biases and how to eliminate or minimise them and the benefits of random allocation are introduced. Qualitative methods including observation, thinking-aloud and interviews are then discussed critically as their non-scientific nature can be seen as methodological weakness; the recommendation from this section is that more than one method (several if practicable) should be used to strengthen the validity of results through correlation.

The development of the SPEL application is documented in Chapter 5 - *SPEL Operation and Interface Design*. The activities delivered by the application are grounded in academic theory. In the absence of existing Child-Computer Interaction (ChiCI) design guidelines for the target age group of 5 to 6 years, it was necessary to develop appropriate guidelines through the development and evaluation of several computer applications for children. The guidelines were used to develop a computer interface appropriate for the target age group of children to ensure that the usability of the application did not affect the results of the RCT that was subsequently carried out to evaluate the effectiveness of the application as a phonemic awareness tutor. A key outcome of this phase of the project was a set of guidelines to inform the design of a usable interface for children aged 5 to 6 years; the significance of the guidelines is discussed in Chapter 5 - *SPEL Operation and Interface Design* and detailed in Section A2 *Design Guidelines Development*.

A pragmatic (within classroom) RCT, carried out to determine the effectiveness of SPEL in developing phonemic awareness is discussed in Chapter 6 - *SPEL Usability and Phonemic Awareness Experiments*. One of the many experimental variables identified in this chapter is the usability of the interface: the work documented in Chapter 5 - *SPEL Operation and Interface Design* enables the control of this variable. The first part of Chapter 6 - *SPEL Usability and Phonemic Awareness Experiments* discusses the qualitative techniques used to evaluate the interface usability leaving the second part of the chapter to document the experimental approach to the quantitative measurement using an RCT to determine the educational effectiveness of SPEL in improving phonemic awareness.

Chapter 2

Computers in the Classroom

Chapter 2 Computers in the Classroom

2.1 Introduction

The focus of this thesis is to determine whether a computer application can be used to develop the phonemic awareness of young children in a classroom setting. Before this area could be tackled, it was necessary to research the background of computers in the classroom to assess the advantages and disadvantages of computer use for literacy, and the preferences and prejudices of teachers, the government and researchers. A literature search was carried out to answer a number of questions which had arisen regarding computers in the classroom; each question is dealt with individually in a question and answer format in the following sections:

2.2 Can computers be used to support learning in the primary classroom?

Historically, there has been a debate concerning the effectiveness of computers in the classroom. At its extremes, the debate has become polarised between those who consider computers to be detrimental to health and learning and those for whom computers can make a key contribution to children's social and intellectual development (Plowman and Stephen, 2003).

In the United States, Cordes and Miller (2000) called for detailed information from medical and commercial bodies researching the physical, mental and developmental hazards computers pose to children. They demanded an immediate moratorium on the further introduction of computers in early childhood and elementary education until the risks were known.

Other critics have argued that it is not the introduction of technology that is the problem, but rather the ways in which it is used. Higgins (2003) reports that although there is some evidence that ICT can assist in pupil learning, there is no evidence that simply using ICT will result in learning. The literature relating to whether or not computers can be used as an effective vehicle for learning in the classroom is largely concerned with preventing what Cuban (2001) calls "a benign addition" to existing teaching practices rather than being used to support new innovative teaching practices. Brindley (2000) points out that multimedia computers provide new and dynamic ways to learn through a range of media including graphics and sound compared to traditional text-based literacy learning. Papert (2004) emphasises aspects of technology that offer the potential for creative problem solving. Plowman and Stephen (2003) suggest that a shift in thinking is necessary if information and communication technology (ICT) is to be effective in education – they need to:

> *"... promote discovery, delight, curiosity, creativity, self-expression and pleasure in learning".*

Technological capabilities such as multimedia appear to have generated a more positive view from the research community with regard to the promising future of computers in the classroom. Underwood et al (2007) report more recent positive attitudes towards computing technology. Ofsted (2005)[1] has reported that technology is increasingly used to enhance the curriculum in imaginative and creative ways that would be impossible without the technology. Christine Gilbert's "2020 Vision: Report of the Teaching and Learning in 2020" (Gilbert, 2007) states that in the future the use of ICT by most children will be the norm and the majority of teachers will have become familiar enough to consider its use as part of traditional teaching practice. Waller (2006) notes that multimedia capabilities of modern computers provide new opportunities for literacy teaching to beginning readers in that:

> *"The combination of image, sound and text could engage young children to attend to textual features and provide support for emergent literacy. Further, images include details and nuances that are more difficult for beginning readers to glean from text".*

2.3 Is ICT useful and effective for reading and spelling in the classroom?

In terms of the use and effectiveness of ICT for reading and spelling in the classroom, a useful source of information is provided by systematic reviews, which select and synthesise high-quality research evidence related to a particular research question. The findings of a systematic review of randomised controlled trials (RCTs) undertaken by Torgerson and Elbourne (2002) evaluating the effect of ICT on spelling, found that the evidence base for the teaching of spelling by using a computer was very weak; a meta-analysis of the studies showed a small, non-statistically significant benefit of computer-assisted literacy learning in spelling. None of the

[1]Office for Standards in Education: the body which assesses the educational standards of schools in England and Wales

RCTs identified in this review were undertaken in the US. The authors note that this is not only important as a reflection of the lack of experimental work in this area in this country, but if the context for the included studies is very different from that in the UK, it may also have implications for the generalisability of the results of the meta-analysis to the UK setting.

The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) is part of the Social Science Research Unit at the Institute of Education, University of London. The EPPI-Centre provides support for those undertaking systematic reviews and provides an evidence database as part of a general move in the UK and elsewhere towards basing policy and professional practice on sound evidence (EPPI, 2010). The EPPI-Centre defines the main elements of a "systematic review" in the following way:

- Explicit and transparent methods are used.
- It is a piece of research following a standard set of stages.
- It is accountable, replicable and updateable.
- There is user involvement to ensure reports are relevant and useful.

The most recent systematic review recorded on the EPPI database evaluating the effectiveness of ICT on literacy learning in English was carried out by Torgerson and Zhu (2003). The review aimed to seek out and analyse studies with the "*most appropriate study design for judging effectiveness: the randomised controlled trial (RCT)*." A randomised controlled trial eliminates selection bias because participants are allocated at random into two or more groups. Allocation of participants randomly ensures that, for the trial population, any differences observed in post-test results can be reliably attributed to the intervention (Torgerson and Torgerson, 2007). For the review of ICT for reading and spelling in the classroom, searches were not carried out for any studies published before 1990; as the authors stated:

> *"the ICT of the 1980s and before was relatively unsophisticated compared with current ICT provision and therefore, trying to inform current ICT policy from studies that used 1980s technology could be misleading."*
>
> (Torgerson and Zhu, 2003)

2,319 potentially relevant reports were identified for the review, but only 12 randomised controlled trials published since 1990 met the inclusion criteria for the evaluation of the effectiveness of ICT for reading and spelling in the classroom. Only studies that had randomly allocated pupils to an ICT or no ICT treatment for the teaching of literacy were included and because the review was an effectiveness review, studies were only included if effect sizes were presented, or enough data was presented to enable the calculation of effect sizes. The review found little evidence of benefit of ICT on literacy learning in the twelve studies, and recommended deferral of further investment in ICT in schools until larger, more rigorously designed randomised trials had been carried out (Torgerson and Torgerson, 2007). The authors also note that the results of the review may be limited in terms of generalisability to the UK since the only studies found to match the inclusion criteria for the review were carried out in the US and point to an urgent need to undertake a pragmatic RCT to evaluate the effectiveness of computer-supported literacy learning in the context of the UK.

Lankshear and Knobel (2003) carried out a review of international research into new technology and early childhood literacy. They found very few mainstream literacy journal articles in this area and almost none concerning literacy in the early years. Wood (2005) and Wild (2009) also claim there is a lack of research relating to the use of ICT to support the early stages of learning to read. It is worthy of note that eleven of the twelve studies included in the review undertaken by Torgerson and Zhu (2003) were carried out with participants aged over 7 years. The only study carried out with children younger than 7 years of age was by Mitchell and Fox (2001) whose 72 participants ranged in age from 5 to 8 years with a mean age of approximately 6.5 years. Their study examined the effectiveness of two computer programs (DaisyQuest and Daisy's castle) on phonological awareness in young children in classrooms in the US. Thirty-six kindergarten and thirty-six first-grade students who demonstrated below grade level performance in reading were randomly assigned to one of three experimental conditions: the DaisyQuest group, a teacher-delivered phonological awareness group or a group using drawing and mathematics software. According to the authors, no differences were found between the DaisyQuest group and the teacher-led group on the total test score.

With regard to research relating to the use of ICT to support the early stages of reading, Underwood (2000) claimed there was a focus on children who had been identified as having specific learning difficulties. This pattern appears to be reflected in the 12 studies included in the review undertaken by Torgerson and Zhu (2003), in which 7 studies used a sample classed as "remedial" as opposed to a sample of mainstream participants with regard to literacy levels at the start of the experiment.

2.4 What do results from evaluations of phonics computer programs indicate?

Underwood (1994) has suggested that ICT applications, even those of the drill and practice type, may be useful in situations where pupils learn within a structured goal-oriented environment and that computer software may be particularly useful in providing a structured practice environment to support phonics reinforcement (Underwood, 2000).

There is a plethora of commercial and free software programs which claim to teach phonics, but according to Slavin's "Best Evidence Synthesis of Effective Reading programs":

> *"As is always true in reviews of educational programs, the largest number of programs by far have never been evaluated in experiments that meet the standards of this review."*
>
> (Slavin et al., 2009)

In 2007, the Department for Education and Skills (DfES) published guidance on how to choose both print-based and computer-supported phonics schemes. They also produced a publisher's self-assessment template to determine how well phonics schemes match their requirements. All the self-assessments which have been submitted are listed on the DfES (2007a) web site. A review of the 37 phonics programs listed on the web site revealed some interesting software assisted phonics tools, such as the Teaching Handwriting, Reading And Spelling Skills Phoneme Machine (THRASS) - a free resource for parents and teachers that uses moving human lips to pronounce the sounds (phonemes) in hundreds of frequently used English words (THRASS, 2009). A search of the literature relating to trials of the phonics schemes available in the UK (including available software) revealed only one connected study with strong evidence of its effectiveness: “Success for all”, which is a non-computer based phonics scheme (Slavin et al., 2009).

The results of RCTs evaluating the educational effectiveness of phonics-based computer applications provided a valuable area for review with regard to locating results pertaining to the effectiveness of specific computer programs. Even though RCTs found in the literature have either been based in the US, been aimed at children older than seven years of age or have been aimed at children with learning difficulties, these studies were examined to find out which phonics-based computer programs were used and how well the software performed. With regard to the ICT trials discussed in Section 2.3, the only study of the educational effectiveness of computerised phonics applications found by the systematic review carried out by Torgerson and Zhu (2003) was the RCT carried out by Mitchell and Fox (2001). Another RCT carried out in the US by Rouse and Krueger (2004) included 374 third, fourth, fifth and sixth grade students (children between eight to eleven years old) who had previously scored in the bottom 20% on the state's reading test. The intervention group used the computer program “FastForward”, which focuses primarily in developing the reading abilities of students; the control group were delivered teacher-led phonics. Even though the authors report that certain aspects of the intervention group children's language skills were slightly improved with the software, Rouse and Krueger (2004) report "*it does not appear that these gains translate ... into actual reading skills*".

More recently, Brooks et al (2006) carried out a randomised controlled trial with children aged eleven to twelve years within a comprehensive school in the North of England. The pupils were randomised to either an intervention group, who received literacy learning delivered via a bespoke phonemic awareness computer application, or to a control group, who received no additional tuition to the usual literacy teaching that is standard practice for the school. No details of the computer application were recorded by the authors. The main purpose of the trial was to look for improvements in spelling scores. The study reported no evidence of a statistically significant benefit on spelling outcome using a computer program for literacy learning. A reduction in reading scores associated with the use of the program was also reported. In the UK, the only RCT found in literature since the study carried out by Brooks et al was a study conducted by Wild (2009), which investigated the use of computers for practising phonological awareness with beginning readers. The study involved 127 children from six primary schools in the UK. The computer application used in the study was the Oxford Reading Tree "Rhyme and Analogy" program, which is based around the theory of analytic phonics. Within each class, children were randomly allocated to one of three groups: 44 children undertook a rhyme and analogy programme using computer software; 43 children followed the same rhyme and analogy programme using comparable paper-based exercises and 40 children used unrelated maths computer games. The study reports that the structured use of literacy software in the year 1 classroom led to greater improvements in the phonological skills of the children who used the computer to support their practising of rhyme and analogy. However, based on the power analysis discussed in Section 6.3.2.1, the small sample size of the groups in six clusters would not enable sufficient statistical power for the experiment to achieve statistical significance.

It is interesting to note that there is no evidence of usability evaluations of the human computer interaction of the applications used in any of the studies discussed in this section. Since poor usability could skew the results against the benefits of computer solutions, it would be prudent for researchers to carry out an evaluation of the usability of an application prior to carrying out a trial to investigate the effectiveness of the applications as a teaching tool. This project gives due consideration to the aspect of usability of computer interfaces used in educational experiments in Chapter 4, Chapter 5, Chapter 6 and Appendix A.

2.5 Are teachers a barrier to the use of computers in the early primary classroom?

A report written for the DfES by Cox et al (2004) claims that different uses of ICT have contributed to some improvements in achievement in English, but the results are inconsistent and restricted by the amount of ICT use and the access to ICT resources in schools. The most commonly reported use of ICT is word processing, although other English-specific software is widely used by some English teachers. The DfES report goes on to say that learners, teachers and managers report that they have a reasonable level of satisfaction with an ICT infrastructure that is reliable, efficient, accessible, affordable and sustainable. However, it is likely that technology will not meet these requirements most of the time.

A BECTA[2] review by Kitchen et al (2006) states that lesson preparation using ICT can save time for teachers by re-using learning material. However, evaluating the material and embedding it into their teaching practice takes time when there is very little time to spare.

Cox et al (2004) claim that teachers' pedagogies have a large effect on pupils' attainment but that insufficient understanding of the scope of an ICT resource leads to inappropriate or superficial uses in the curriculum. Ofsted (2006) collected evidence from 30 colleges and 13 universities which showed that student teachers did not receive the level of support needed to develop their literacy, numeracy and ICT skills. This becomes problematic when the teacher is unable to effectively direct the pupils in the use of ICT. Kitchen et al (2006) claim the main uses of ICT across the curriculum are word processing, Internet access and presentations. Waller (2006) suggests that, even though new technology such as interactive whiteboards have now been introduced into early childhood, many early years educators are unsure of how to make use of the technology. Shenton and Pagett (2007) carried out an observational study of the use of interactive whiteboards in six English primary classrooms. They report that in most of the lessons observed, it was usually the teachers who used the interactive whiteboard controls and that the board was used primarily as the teacher's tool with little or no opportunity for children to interact with the technology. Related further work by Kitchen et al (2007) shows the need for teacher development in using classroom technology with pupils: most of the teachers could use the Internet but about 75% of the teachers canvassed expressed a need for training in particular software packages.

2.6 Is gender a barrier to the use of computers in the early primary classroom?

The relationship between gender and attainment using a computer as a learning vehicle needs to be reviewed to establish at this early stage in the research whether gender is known to be significant in the effectiveness of computer–based instructional programmes.

[2] British Educational Communications and Technology Agency was an agency of Department for Education and Skills in the UK.

The UK DFES Standards Site states that performance differences relating to gender are *"a matter of national concern"* (DfES, 2010).

With regard to performance in English, the UK National Literacy Trust state:

> *"Although the headlines exaggerate the problem, there is consistent evidence that boys' achievement lags behind that of girls – a trend that is international. In England, the discrepancy is particularly evident in English, where statistics show that boys' performance is lower than girls' in all literacy related tasks and tests, and a significant percentage of boys is not attracted to reading."*

(Clark and Akerman, 2008)

With regard to performance related to ICT, Volman et al (2005) reported findings from a study in seven schools (primary and secondary). Data were collected on participation, ICT skills and learning results, ICT attitudes and the learning approach of pupils. A total of 213 pupils completed a questionnaire and interviews were held with 48 pupils and 12 teachers. The researchers report small gender differences in primary schools.

With regard to computer use outside school, a DfES study of home use of ICT by Valentine et al (2005) reported gender implications from a study of how children's home use of computers affected their attainment at school. A gendered pattern of ICT use was found in this study as early as year 2: boys used computers more at home for "fun" and girls used ICT for educational purposes. Valentine et al note that this has implications for the gender educational gap as the research also showed that high levels of leisure use of ICT were correlated with a negative impact on educational attainment.

2.7 Is age a barrier to computer-based learning in the early years?

Given government initiatives in different countries to introduce ICT at progressively earlier stages of education, Plowman and Stephen (2003) note that there is a lively debate by parents, practitioners and researchers on the desirability of such policies.

Healy (1998) adopts a negative stance towards the impact of computer use on young children's learning. She argues that the use of computers is damaging to young children's development and therefore to their learning because the early years are a 'busy time for the brain' and using computers before the age of seven interferes with other important developmental tasks. She emphasises the need of children for human support and verbal interaction in the early years and concludes that computers are an inappropriate learning tool for children below the age of about 7 years. This view is backed up by Haugland (1999), who states that children learn through their bodies so computers are not developmentally appropriate. Blakemore et al (2004) reviewed developments in neuroscience and reported that even though much early learning seems to be automatic, children require a rich and stimulating environment in which to learn and that social interaction with others seems to be key.

Rather than asking at what age technology should be introduced to children, Van Scoter et al (2001) claim it would be more useful to ask, *"What are appropriate and meaningful uses of technology with children?"* Kinder (1991) believes that children seem to be able to move between certain types of media, such as televisions and computer games with ease. In this sense, the technology itself may be less important than how it is incorporated into learning environments. Following a study in 2007, BECTA, who have responsibility for supporting technology for learning in schools, reported positive outcomes if ICT was applied in a particular way:

> *"Where ICT has become a regular part of the classroom experience, there is evidence of positive impact on learning and pupil performance. Various studies have found evidence that the visual nature of some technologies, particularly animations, simulations and moving imagery, engaged learners and enhanced conceptual understanding."*
>
> (Condie et al., 2007)

2.8 Conclusions

Although there are opposing opinions about computer use by young children, when looking at the research literature holistically, it can be seen that most of the views are valid in their own context. The following conclusions have helped to formulate an overview of the debates involving the area of ICT for young children in the primary classroom and helped to identify the gaps in research within this area:

- Section 2.2 discusses historical concerns regarding the effects of computers on the health and learning of children. Multimedia computers, however, offer new and dynamic ways to learn through a range of media including graphics and sound compared to traditional text-based literacy learning.

- The lack of quantitative evidence regarding the usefulness and effectiveness of ICT for reading and spelling in the classroom is identified in Section 2.3. This thesis contributes to the quantitative evidence in this area in Chapter 6.

- The shortage of randomised controlled trials to evaluate the effectiveness of computerised phonics programs in UK primary classrooms, particularly for children younger than seven years of age in main stream education is discussed in Section 2.4. This thesis contributes to the quantitative evidence in this area in Chapter 6.

- Section 2.5 discusses how teachers and developers tend to "shoehorn" existing teaching practice into the computer domain. This may be true but not necessarily always a bad thing as some subject areas lend themselves particularly well to computerisation: a computer can be used as a learning tool as well as a learning environment if the facilities offered by the computer are appropriate for specific activities. For example, areas of teaching that rely on sound, graphics and/or speech recognition can lend themselves well to some learning tasks. Technology designers and teachers should identify areas where such facilities are fundamental to the teaching of the subject and as such could provide useful teaching tools for one or both of the following reasons: to lighten the teaching overhead, thus freeing up valuable teacher time to spend on tasks which require a more socially interactive teaching element; or to provide a more modern and better way of presenting information, as a pen and paper for example, cannot speak, listen or dynamically display relevant graphics.

Examples of such areas are: speech recognition to help a child practise reading and to teach phonics. The traditional teaching process would involve a teacher listening to a child read and providing timely feedback, so the speech recognition and the audio output capabilities of the computer are a fundamental part of the teaching process and the computer is well suited to this task. Similarly, phonics is an area in which the facilities offered by multimedia, such as the graphics, sounds and interactive feedback are fundamental to the teaching process of mapping phonetic sounds to graphical symbols.

There is evidence that many teachers lack support and/or training in the use of ICT. To counteract these problems, work needs to be carried out in the area of Child-Computer Interaction to ensure the computer applications are very intuitive and simple to use so will minimise the effect of insufficient training; work undertaken to develop this area is this is the subject of Chapter 5 and Chapter 6, with further detail in Appendix A.

It appears that there is some reluctance, even resistance, to the use of computers in the classroom. However, it is really the lack of solid research which quantifies the benefits or identifies the hazards which is lacking. Although the literature apparently expresses many diverse opinions, a suggested summary of the underlying message from the bulk of the literature reviewed is:

Do not put great emphasis on something new without showing that its use will safely enhance the learning experience rather than detract from it or present developmental hazards.

The introduction of ICT requires careful consideration of change management; not all teachers will want to learn “new tricks” so the management of such changes is necessary to encourage those people to adapt. Others feel they have no time to teach themselves and require formal training to get the best from the technology and appear competent to their students. If technology is to become firmly integrated into the teaching fabric, it needs to be reliable: having the network or server fail during a carefully prepared class or during a computerised test is unlikely to win over the “talk and chalk” diehards.

Research discussed in Section 2.6 indicates a gender difference in computer use in primary schools, so it needs to be considered a possible bias factor in this research. In the experiment covered in Chapter 6, gender is considered to be a variable which needs to be controlled.

Age is a variable that also needs controlling because older children may perform better than younger children. In year 1 of primary school, even though the children are in the same year; children’s ages range from 5 to 6 years so a child could be 20% older than another at the extreme.

There is no significant evidence that software products should not be developed for young children; research discussed in Section 2.7 indicates that there is opportunity to enhance the learning experience of young children using the visual and audible features of modern multimedia technology.

Based on discussion of the area of ICT for young children in the primary classroom in this chapter, there seems to be a case for developing software to assist with the learning of phonemic awareness. However, usability of the computer interface is critical to ensure that young children can use it without tying up teacher time or becoming confused or frustrated with its operation. The product ought to be evaluated formally using an RCT to provide proper evidence of its effectiveness as a learning tool. The computer application designed for this study aims to develop phonemic awareness in young children which is a key component of phonics. A discussion of phonics is therefore the subject of the next chapter: *Phonics in UK classrooms: the debate*.

Chapter 3

Phonics in UK classrooms: the debate

Chapter 3 Phonics in UK classrooms: the debate

3.1 Introduction

This chapter explains the phonics approach to teaching reading in the early primary years in the UK. Teaching strategies for literacy have been vigorously researched and debated over the years. The focus of this chapter is to critically review this research in order to utilise and develop it in subsequent chapters of this thesis. The chapter concludes that synthetic phonics is the UK's primary approach to teaching reading in the early years. This conclusion underpins the work in subsequent chapters which discusses the development and evaluation of a computerised phonics tutor.

3.2 Phonics – an overview

> *"A phoneme is a distinctive speech sound which will make a difference to the meaning of a word. For example, the initial phonemes in bat, pat are /b, p/. A grapheme is a letter or combination of letters used to spell a phoneme, for example the letters <p, sh> spelling the phonemes /p, ʃ/ in push."*
>
> (Torgerson et al., 2006)

> *"Spoken English, in most of the accents found in Britain, has about 44 phonemes."*
>
> (Brooks, 2003)

There is a one-to-many relationship between phonemes and graphemes. There is also a one-to-many relationship mapping of some graphemes to phonemes. Phonics is a method of teaching reading and writing at a word level using the sound↔symbol (or phoneme↔grapheme) correspondences of a language. Before phonics can be used, a child must have or be able to gain phonemic awareness, which is the awareness that spoken words are made up of phonetic sounds and written words are made up of grapheme symbols and that spoken words and written words are linked by the sound↔symbol correspondences of phonemes and graphemes. There are two prevalent approaches to phonics: synthetic phonics and analytic phonics.

3.2.1 **Synthetic phonics**

The synthetic phonics approach teaches children that the sounds associated with graphemes make up words. By learning how to "blend" phonemes into words (synthesise the word) the beginner-reader can sound-out written words by mapping the graphemes in the word to the appropriate phonemes, then by sounding-out the phonemes sequentially they can say the word. There is a complementary skill, "segmenting", which enables a child to decode written words into their constituent phonemes. The word is decomposed into its grapheme set then each grapheme is sounded out using its corresponding phoneme to form the spoken representation of each part of the word. Segmenting is the decomposition of a word into its spelling such as CAT which is segmented into the graphemes <c> <a> <t> and sounded out individually as "kuh-a-tuh" (where the "uh" is not actually sounded).

The techniques of blending and segmenting can be used in isolation or together. For example, to read a word a beginner-reader would segment the word into its phonemes and then blend the phonemes into the spoken word. If a child wishes to spell a word they hear or already know, they would segment the word into its individual phonemes, map them to the appropriate graphemes then write, point at, use magnetic graphemes or other means to build up the written word.

3.2.2 **Analytic phonics**

In analytic phonics (also known as implicit phonics), the phonemes associated with the written graphemes are not pronounced in isolation. Children are taught to recognise the beginning and ending sounds of words without breaking these down into individual phonemes. Typically this approach groups words with common starting or ending sounds and the words are learnt at this group level. For example, *pet, park, push* and *pen* all begin with the phoneme */p/.* This common starting sound is referred to as the ***onset*** of the word.

Similarly, words are "analysed" to determine groupings of similar ending sounds. For example, *bat, cat, fat, hat* and *mat* all end with the "at" sound. This common ending sound is referred to as the ***rime***. By learning groups of onsets and rimes, children can concatenate different onsets with different rimes to make new words.

3.3 The development of Phonics teaching

Figure 3-1 illustrates the timeline identifying key milestones in the evolution of reading teaching in the United Kingdom. As the approach to teaching literacy has been fiercely debated for a number of years, particularly in the area of phonics, the timeline is a summary of key events in that period. The remainder of this section discusses the timeline in more detail with particular emphasis on the development of the teaching of phonics since the implementation of the National Literacy Strategy.

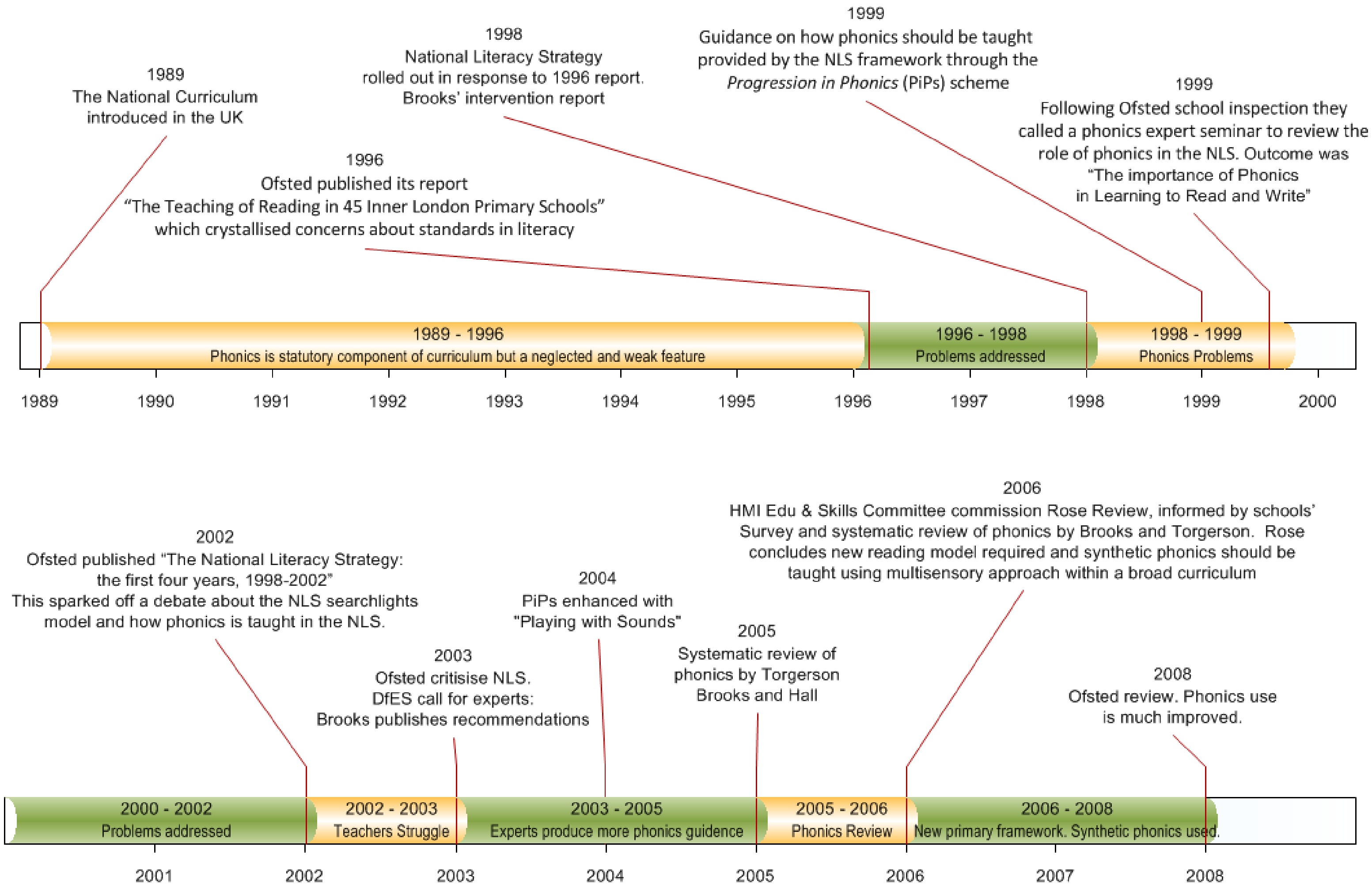

Figure 3-1: Phonics' development timeline

Although all phonics methods require sound↔symbol mapping, the subtle differences in the use or granularity of the sounds is still widely debated. Good quality research in the area of literacy, of which phonics is a part, is essential to improve the standards of reading and writing. The results of such research can be used to inform the government which can then make informed decisions on the teaching approach to adopt in schools. Over the years since the introduction of the National Literacy Strategy, the body of research has been distilled into government reports which set a common teaching approach. These reports and the research within them are discussed in this section and inform the chapter's conclusions.

3.3.1 **1989 - 1998**

Views of the role of phonics in the teaching of literacy have changed radically over the years. Marilyn Adams, a powerful influence in confirming the value of phonics, reported:

> *"Perhaps the most influential arguments for teaching phonics are based on studies comparing the relative effectiveness of different approaches to teaching beginning reading. Collectively these studies suggest, with impressive consistency, that programs including systematic instruction on letter-to-sound correspondences lead to higher achievement, at least in the early grades and especially for slower or economically disadvantaged students."*
>
> (Adams, 1990)

The National Curriculum was introduced in the UK in 1989 incorporating 10 subjects and national testing. According to Sir Jim Rose (the former Director of Inspection at Ofsted), despite phonics being a compulsory component of the National Curriculum (NC), over the first nine years of the NC, it was often either neglected or not treated seriously (Rose, 2006). However, the introduction of the National Literacy Strategy in 1998 improved this situation because the government engaged schools in developing a structured approach to literacy including how phonics should be taught (Rose, 2006).

3.3.2 **1998**

Following a pilot project in 1996, which involved schools in 14 Local Education Authorities, the National Literacy Strategy (NLS) was introduced to all primary schools in England in September 1998. The strategy required 1 hour of literacy to be taught each day; this was known as the "literacy hour". The main purpose of the literacy hour was to raise literacy standards by classroom teaching which focused on reading and writing skills for 60 minutes each day. However, it was crucial that teachers also developed their own subject knowledge at the same time in order to raise standards.

The National Literacy Strategy advocated the "searchlights" model for reading (Figure 3-2). The searchlights model promoted reading in a broad sense in that teachers could encourage children to "switch on" any of the four searchlights of: phonics; contextual knowledge; grammatical knowledge; word recognition and graphic knowledge, during the reading and writing process. None of the strategies was given priority over the others in this model and commentary on the model provided in the NLS suggests that children who could use more than one of these strategies at once would be overall better readers:

> *"The more 'searchlights' that are switched on, the less critical it is if one of them fails."*
>
> (Education and Skills Committee House of Commons, 2005)

According to Beard:

> *"The need for such an integration is acknowledged in the structure of the Literacy Hour which ensures that text level, sentence level and word level objectives are consistently addressed and cross-referenced. Drawing upon these different sources of information in fluent reading is referred to in the Framework as using the full range of 'searchlights' in tackling texts."*
>
> (Beard, 1999)

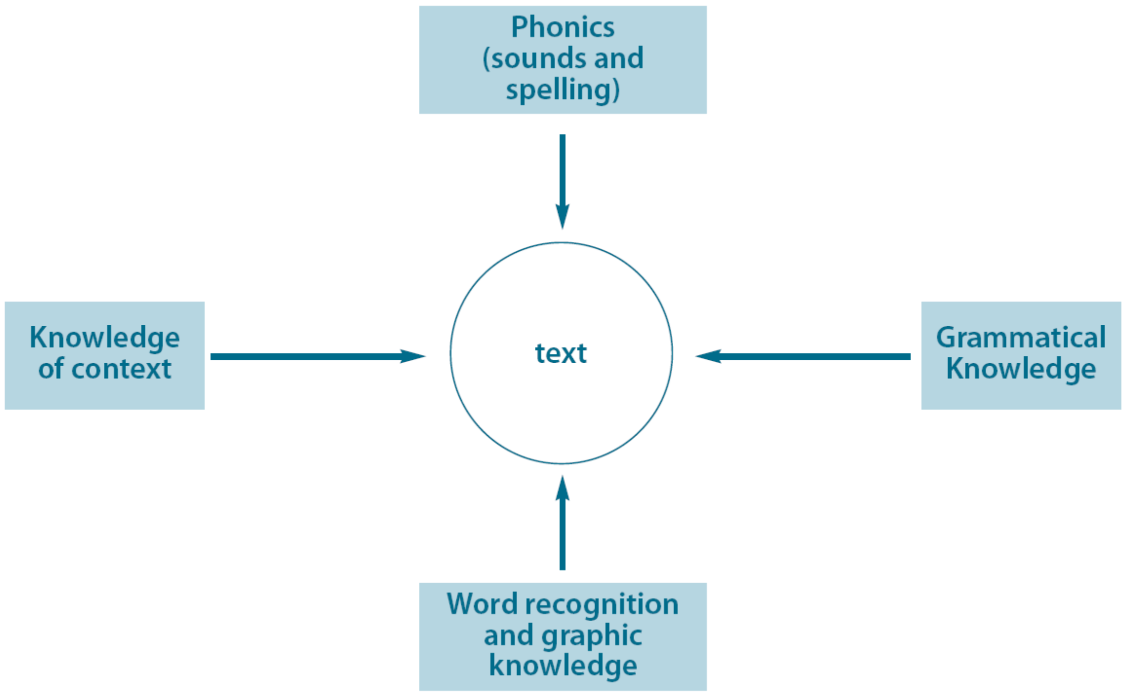

Figure 3-2: The NLS Searchlights Model

(Rose, 2006)

3.3.3 **1999**

In 1999, after an evaluation by HMI (Ofsted, 1998) of the first year of the NLS, it became clear that the teaching of phonics was not well understood by teachers. To address this problem, guidance on how phonics should be taught was provided by the NLS framework through the *Progression in Phonics* (PiPs) scheme. This scheme set out a programme for children of 15 minutes each day over 4 terms (summer reception to summer year 1) to support the learning of basic decoding and encoding skills. A team from the Ontario Institute for studies in Education at the University of Toronto, who specialised in the area of large-scale educational reform, was commissioned by the DfES to carry out an independent "*evaluation of the implementation of the National Literacy and Numeracy strategies*" to complement the HMI evaluation.

The first Ontario report listed in its conclusions:

> *"The Strategies are off to an impressive start, but if they are to be sustained, local educators must continue to increase their understanding of literacy…and must also feel a sense of ownership."*
>
> (Earl et al., 2000)

A review by Brooks et al (1998) looked at the effectiveness of schemes that had been developed to improve the reading attainment of slow readers (excluding dyslexics) in years 1 to 4 in the UK. The report recognised that the roll-out of the National Literacy Strategy in Q3 of 1998 would put the onus upon teachers and educationalists to make important choices with regard to their selection of intervention schemes. It also recognised the need to address the issue that 20% of children in the UK had not achieved level 2 in reading by the time they were seven years old. Brooks et al (1998) reviewed the research publications in this area and found a wide range of quality which led to them discount most work as it either provided no quantitative data or provided data which were unusable due to flawed processes.

Brooks and his colleagues noted that data had been found to be missing even from the reports of quite large-scale independently-funded evaluations. The evaluation subsequently cites a very useful list of the minimum information that should be provided by researchers reporting on studies such as these. These recommendations have been cited and used as guidelines in Chapter 6 of this thesis which covers a classroom experiment on the effectiveness of a custom-built phonemic awareness tutor.

3.3.4 **2002**

In 2002, in the report "*TheNational Literacy Strategy: the first four years, 1998-2002"*, Ofsted acknowledged the improvement in spelling ability was almost certainly due to the teaching of phonics at Key Stage 1. However, it was found that the improved application of phonics teaching had not permeated through to the reception year.

In the same report, Ofsted criticised the searchlight model's representation of phonics in the reading process as it was not clear what the intensity of each searchlight should be at different reading progression points. The report concluded that there had been a shift towards word-level work which detracts from the need to blend sounds together using grapheme to sound correspondences (Ofsted., 2002).

3.3.5 **2003**

The weaknesses in the teaching of phonics identified in the 2002 Ofsted report resulted in the DfES calling for a phonics expert conference to be held in London on 17 March 2003. The conference, chaired by Professor Greg Brooks, was attended by researchers, practitioners, members of the National Primary Strategy reference group, officials from the DfES, the Qualifications and Curriculum Authority (QCA) and Ofsted. The findings from the conference resulted in the report "Sound Sense: the phonics element of the National Literacy Strategy. A report to the Department for Education and Skills" (Brooks, 2003). The report addressed the question:

> *"To what extent, and in what ways, does the phonics element of the National Literacy Strategy need modifying?"*

(Brooks, 2003)

The report found that a major redirection of the phonics element of the NLS was not necessary. However, some revisions to the phonics element of the NLS were required and there was a need for some focused research in specific areas of phonics. With regard to the searchlights model, Brooks explained that the model was being interpreted in a simplistic manner which gave an equal focus on each searchlight; this is not the way the model should work – the brightness of each searchlight should vary depending on the stage of learning. Brooks goes on to explain that fluent readers have built up a good sight vocabulary but revert to decoding unfamiliar words rather than guessing the words based on its context, whereas poor readers try to guess words based on their context and often get them wrong; they need to improve their decoding skills. Findings from the Brooks (2003) report of particular relevance to this thesis are now discussed in the form of answers to key questions:

Is phonics necessary?

Phonics is without doubt a key reading development skill. The (US) National Reading Panel (Ehri et al., 2001, National Reading Panel, 2000) showed that children who were taught using systematic phonics progressed better in reading and spelling than children taught using a less organised approach or no phonics at all.

Should synthetic or analytic phonics be used?

Experimental evidence indicates similar success rate from both. However, there is little empirical evidence to clearly compare the effectiveness of both. The approach proposed in the National Literacy Strategy is synthetic phonics.

Can synthetic phonics be used for both reading and spelling?

Synthetic phonics refers to an approach to the teaching of reading in which the phonemes associated with particular graphemes are pronounced in isolation and blended together (synthesised). Synthetic phonics for spelling involves analysis, namely the segmentation of spoken words into phonemes.

Do beginning readers need to be taught sight words?

If a word is not decipherable using the phonics approach then some words will need to be learnt by sight. A small initial sight vocabulary should be taught, but this does not need to comprise the whole list of the 100 most frequent words. The phonically regular words within that list should be taught phonically.

When and at what pace should phonics be taught?

Phonics teaching can start in the reception year then progressed in year 1. It should be taught quickly and systematically. It should not take a formal approach but be interactive, active, lively and fun.

In which order should phonics be taught?

A sensible approach is to learn letters and their sounds as this quickly enables many Consonant-Vowel-Consonant (CVC) words to be handled. Johnston and Watson (2005) and Augur and Briggs (1992) suggest starting with **<*s, a, t, p, i, n*>**.

How soon should all-through-the-word phonics be introduced?

As soon as the small starting set of letters has been learnt, grapheme-phoneme mapping needs to be learnt to build up whole words. As all English words contain at least one vowel, vowels need to be introduced early on in this more advanced phase.

3.3.6 **2004**

In 2004, the DfES published a phonics supplement to "Progression in Phonics" (PiPs) called "Playing with Sounds". Although *Playing with Sounds* had been designed to provide more detailed guidance on how to teach phonics in the classroom, Lesley Drake and Debbie Hepplewhite from the Reading Reform Foundation (RRF) (2004) voiced concerns about the suggested slower pace of phonics teaching and the order of sounding out words from first-last-middle; it is illogical, particularly when teaching programmes using first-middle-last sounds have been successful and they found no empirical evidence to support this change.

3.3.7 **2005**

In 2005, to bring the National Literacy Strategy in line with research and findings since it was published in 1998, the HMI Education and Skills Committee (ESC) published the report, "Teaching children to read". The report acknowledges the improvement in literacy standards in primary schools and the desire to build on that success by keeping up to date with current research and current practice. A key recommendation of the report is that the government should review the NLS and the DfES should commission a large scale study to compare the NLS with the "phonics fast and first" approaches. At this point in the evolution the debate has shifted from, "Should phonics be taught?" to "How should phonics be taught?" (Education and Skills Committee House of Commons, 2005).

Rose was asked to conduct an independent review of the full range of best practice in the teaching of early reading and the strategies that best support children who have fallen behind in reading. Rose was to do this through examination of the available evidence and engagement with the teaching profession and education experts. In June 2005, an Education and Skills Committee report explained that Carole Torgerson and Greg Brooks had been commissioned to carry out an independent review of phonics teaching and its application. Rose would be able to use the findings of Torgerson's and Brooks' analysis to inform any modifications to the NLS. To ensure the future built on the success of the current implementation of phonics teaching, a pilot study was commissioned; this engaged 200 primary schools and was based on the Primary National Strategy's "Playing with Sounds" programme (Education and Skills Select Committee, 2005).

Rose conducted a full review of the research findings, carried out a full analysis of the evidence already available of what is working in schools and commissioned Ofsted to undertake some rapid review work to observe the features of "best practice" in synthetic phonics and in using the National Literacy framework for teaching.

3.3.8 **2006**

A systematic review of experimental research on the use of phonics instruction in the teaching of reading and spelling commissioned by the DfES undertaken by Torgerson et al was published in 2006:

> *"This review built on a systematic review conducted in the United States by the National Reading Panel's phonics subgroup (Ehri et al., 2001), which concluded that systematic phonics teaching helped children learn to read better than all forms of control group teaching."*

(Torgerson et al., 2006)

From the research, Torgerson et al identified twenty RCTs and of those, only Johnston and Watson (2004, experiment 2) was carried out in the UK. A later experiment (Johnston and Watson, 2005) made a significant impact on literacy policy in England. The results were cited by the Education and Skills Committee House of Commons (2005) as one of the reasons that the government should undertake an immediate review of the National Literacy Strategy. The Rose report recommended that all English children should be taught to use synthetic phonics as a primary approach to learning to read; results from the Clackmannanshire study were influential in this recommendation (Ellis, 2007).

The key conclusions of the systematic review (Torgerson et al., 2006) relevant to this thesis are:

- Systematic phonics instruction within a broad literacy curriculum was found to have a statistically significant positive effect on reading accuracy.

- There was no statistically significant difference between the effectiveness of systematic phonics instruction for reading accuracy for normally-developing children and for children at risk of reading failure.

Both of these conclusions provided some support for the findings of a systematic review published in the United States in 2001 (Ehri et al., 2001).

Note: the weight of evidence for both of these findings was moderate (there were 12 randomised controlled trials included in the analysis).

Interestingly, no statistically significant difference in effectiveness was found between synthetic phonics instruction and analytic phonics instruction so this did little to resolve the "analytic or synthetic" debate. However, there were only three randomized controlled trials on this topic, so no informed decision could be made regarding one method or the other without further work.

The Rose Review (2006) commissioned by the DfES concentrated on good practice in the teaching of reading, including good practice in the use of phonics. The review drew upon three main sources of information: the findings of research and inspection; wide-ranging consultation, including practitioners, teachers, trainers, resource providers and policy makers; visits to schools and training events. The report uses studies in the area of word recognition and comprehension to justify the replacement of the searchlight model with that of the "simple model of reading" illustrated in Figure 3-3.

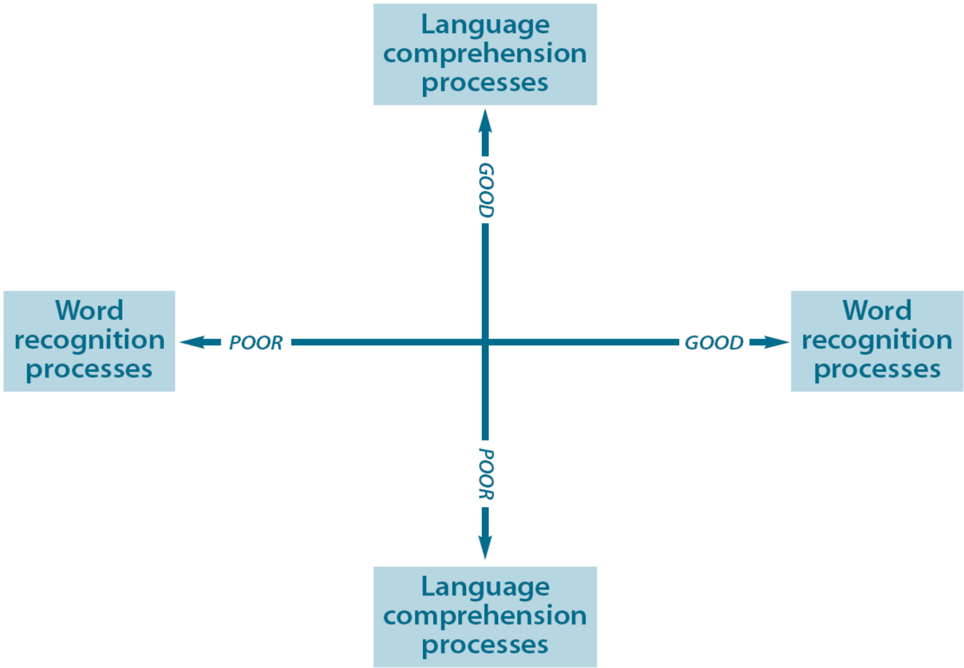

Figure 3-3: Simplified reading model

(Rose, 2006)

Rose based the model on supporting research literature explaining that beginner-readers need to set up processes for identifying letters, as words are made up of letters:

> *"However, it is evident from the research literature that the balance of learning needs across the two dimensions changes as children become more fluent and automatic readers of words: that is, establishing the cognitive processes that underlie fluent automatic word reading is a time limited task, and involves acquiring and practising certain skills, whereas developing the abilities necessary to understanding and appreciating written texts in different content areas and literary genres continues throughout the lifespan."*
>
> (Rose, 2006)

3.3.9 **2007**

In 2007, Brooks was commissioned to update his report "What Works for Slow Readers?" (Brooks et al., 1998), previously revised in 2002 (Brooks, 2002b). In the 2007 revision, Brooks states that the evaluation was restricted to schemes used and evaluated in the UK because schemes used elsewhere in the world may not necessarily work in the UK (Brooks, 2007b).

The requirement for the study came from statistics which revealed in 1998 that 19% of children at key stage 1 and 7% at key stage 2 were experiencing literacy difficulties and despite endless efforts to improve standards, the number experiencing literacy difficulties in 2006 was still 16% at key stage 1 and 6% at key stage 2; a significant number of children will therefore struggle at Key Stage 3 and beyond.

With regard to ICT studies, Brooks found that gains could be made if the technology was used correctly and supported by teachers. However, where this support was unavailable, the children tended to flounder. The report also includes a useful set of recommendations for researchers and developers designed to validate literacy programmes.

In this year, the DfES published guidance on how to choose both print-based and computer-supported phonics schemes. They also produced a publisher's self-assessment template to determine how well computer applications match the requirements set by the DfES; the template is available from the DfES website (2008).

The UK Government published another set of guidance documents on the teaching of phonics called "Letters and Sounds"; this replaced "PiPs" and "Playing with Sounds". "Letters and Sounds" is based around the findings from the Rose Review which acknowledges the importance of early phonics teaching. It goes on to suggest that a "multi-sensory" approach would enhance the child's phonic knowledge and skill.

"Letters and Sounds" is a six phase approach: in phase one, children are encouraged to develop their aural language skills; in phases two to six they carry out high quality phonics work. The guidance suggests that phonics is best taught in short discrete daily sessions of around 20 minutes, but supported with opportunities to use and apply their knowledge and skills of phonics throughout the day. Phase two starts with the established approach to learning a set of letters (s,a,t,p,i,n) with an emphasis on multi-sensory activity; at the end of phase two the majority of children should have mastered decoding print. The processes of segmenting and blending for reading and spelling need to be made fun and easy to understand and use. Children need to compose words by manipulating letters even though they may not yet be able to write them; magnetic boards provide ready formed letters for the children to experiment with. During the early phases, children need to build up a mental database of the grapheme-phoneme mapping (DfES, 2007b).

3.4 Conclusions

A literature search in the development of the UK curriculum in literacy teaching illustrates an iterative evolutionary development. Teachers were initially expected to be able to deliver the programme and impact on children's standards of attainment at a time when the research evidence was either not available or incomplete.

The teaching of phonics has been particularly troublesome. It is clear that phonics has been the preferred approach to the initial phases of reading and writing for some time. However, it was not initially being used effectively by teachers and the phonics technique of choice was hotly debated.

Phonics has remained an essential component of early reading over the years as is evidenced by the reviews and modifications to the teaching processes with synthetic phonics surfacing as the technique of choice in UK classrooms today.

Recent recommendations suggest that synthetic phonics teaching should use a multi-sensory approach where possible and teachers should ensure that their classroom delivery of phonics provides consistent enunciation as correct and consistent pronunciation of the phonics sounds is required for the technique to work properly. A computer system with pre-recorded speech, for example, will deliver a consistent enunciation and correct pronunciation if recorded properly. However, no work was uncovered in the literature review which examined the effects of teachers with strong regional accents delivering a speech-related component of the curriculum (where pronunciation of phonic sounds and words is important); the effects of regional accents on phonics teaching could be an interesting area of further work.

A literature search has been carried out to inform the development of a synthetic phonics application. The application, SPEL, has been developed as a prototype to determine the effectiveness of teaching phonemic awareness using a computer program. The application and its effectiveness are subjects of later chapters.

Examples of how some of the main conclusions from the research detailed in this chapter have informed and validated the design of SPEL are summarised here:

- ***Systematic phonics instruction within a broad literacy curriculum has been found to have a statistically significant positive effect on reading accuracy*;**
  SPEL provides a systematic and consistent platform for learning by the very nature of it being a machine.

- ***There was no statistically significant difference found between the effectiveness of systematic phonics instruction for reading accuracy for normally-developing children and for children at risk of reading failure;***
  a computer program with a suitably simple and intuitive interface could conceivably be used to teach fast and slow readers; key aims of the SPEL interface are simplicity and usability. SPEL does not "bully" the child through the activity but allows them to explore and work at their own pace.

- ***Experimental evidence does not tend to show that synthetic phonics produces better results than analytic phonics. However, synthetic phonics is embodied in and advocated by the UK curriculum*;**
  SPEL is based on the synthetic phonics approach to align with curriculum requirements. An experiment in a classroom setting was carried out to determine its effectiveness as a phonemic awareness tutor (Chapter 6).

- ***Sight vocabulary becomes a database from which children can infer more sophisticated complex and conditional phonic rules*;**

  SPEL enables a child to encode and decode words whilst hearing the individual sounds that make up a word - then hear the sound of the complete word; seeing the phonics alphabet and hearing the pronunciation of each grapheme will assist in memorising the sound↔symbol mapping of phonemes and complete words; this should enable and encourage children to develop their own words. SPEL does not attempt to teach sight vocabulary as it is designed primarily as a phonemic awareness tutor.

- ***The 'simple model' of literacy learning has replaced the four strand searchlight model in the NLS and symbolises the relationship between phonics and comprehension*;**

  phonics teaching takes priority for beginning readers, but the balance of learning needs across the dimensions changes as children become more fluent and automatic readers of words. This means that phonics teaching is a time limited task. SPEL can be used as a tool in the development of sound↔symbol mapping; such a system however should not become a substitute for interaction with teachers, other children and comprehension work.

- ***The teaching of phonics should be systematic, quick and early, as opposed to incidental, slow or late*;**

  SPEL systematically teaches sound↔symbol correspondence in the time frame and to appropriate age groups of children as specified by the Letters and Sounds DfES recommended phonics scheme.

- ***A logical order in which to teach the phonic alphabet would be to pick letter-sounds which build up rapidly into a set which provides a reasonably sized vocabulary of regular CVC words. The six letters <s, a, t, p, i, n> proposed by (Johnston and Watson, 2005) and (Augur and Briggs, 1992), do just that*;**

  SPEL is based around the Letters and Sounds DfES recommended phonics scheme, which begins with the teaching of the six letters <s, a, t, p, i, n>.

- ***Sounds should be learnt in the order First-Middle-Last***

  The order in which the sounds are learnt is important; the SPEL application uses the logical approach of first-middle-last sounds based on work carried out later in the NLS evolution.

- ***Gains could be made if the technology was used correctly and supported by teachers***

  This finding is particularly pertinent to this thesis as usability is considered key to the success of the application. The child should always know or be able to reasonably determine what is required of them without the need to guess or continually ask for help from the teacher.

- ***Phonics should be taught with emphasis on multi-sensory activities***

  One benefit that the computer program has over magnetic letters on whiteboards is the ability of the computer to make the phonic and complete word sounds. This enables children to map the sound they expected to hear from their word building exercise with the actual sound of the finished word. However, magnetic letters draw on the ***kinaesthetic learning skills*** in a way that using a mouse doesn't. Magnetic letters may encourage sharing and interaction and can be used to label physical objects; a computer program is not proposed as a replacement for a magnetic board or a teacher but simply another learning tool with some additional or alternative benefits.

- ***Phonics teaching can start in the reception year then progressed in year 1***

  The pace and content of SPEL follows the Letters and Sounds phonics scheme. SPEL is designed for children aged between 5 to 6 years (year 1 of primary school) and as such it was necessary to design and evaluate an interface specifically for this age group as they are at best beginner-readers.

This chapter has discussed the development of phonics up to the current time. As this has been evolutionary and convergent, it is unlikely to change fundamentally in the near future so a phonics tutor which satisfies current teaching requirements could be a very useful classroom tool. However, for the tool to be effective, like any good tool it should be simple to use and get the job done with the minimum of fuss and frills. A hammer, for example, is a very effective tool; it is simple in design, it is easy and intuitive to use, it has no fuss nor frills to get in the way of its use – the author believes an educational tool should emulate this simplicity.

In addition to findings from literature reviews and validating the design of SPEL against the research findings, recommendations on randomised controlled trials (Brooks et al., 1998, Brooks, 2002b, Brooks, 2007b) informed the RCT experiment on the SPEL application; this is detailed in Chapter 6.

Simply designing a computer application for young children is not enough - it needs to be evaluated for usability and effectiveness in its goals. For such measurements to be meaningful they need to use reliable evaluation methods; such methods are the subject of the next chapter: *Research and Evaluation Methods*.

Chapter 4

Research and Evaluation Methods

Chapter 4 Research and Evaluation Methods

4.1 Introduction

To determine the educational effectiveness of the System for Phonic Early Learning (SPEL) application it is necessary to design an appropriate evaluation approach using a well-defined process such that the results will be meaningful and valid. This chapter discusses the methods used in the evaluations and experiments carried out for this project.

Quantitative and qualitative research methods are discussed in this section and their appropriateness for each part of the project evaluation. Quantitative research methods are characterised by numerical analysis, whereas qualitative methods are characterised by the use of narrative accounts (Clissett, 2008). Quantitative research methods produce factual, reliable, and generalisable data, whereas qualitative methods generate rich, detailed, valid process data (Steckler et al., 1992). Quantitative research involves the collection and analysis of numerical data whereas qualitative research involves analysis of data such as words and pictures from interviews, observations and video recordings for example.

4.2 Quantitative research method

There are a large number of research methods so only the techniques chosen for this project will be covered. An overview of the general experimental method is discussed here leaving detail specific to the implementation of the experiment until Chapter 6 - *SPEL Usability and Phonemic Awareness Experiments*.

Boruch et al refer to the Randomized Control Trial (RCT) as the "gold standard" method for testing effects of different interventions in many fields, including social programmes, since the 1970s (Boruch et al., 2002). Using this method of elegant design, subjects are allocated to two or more groups randomly and then either exposed to an intervention or to a control or comparison condition, which can either be an alternative intervention or no treatment control (Torgerson, 2009). Because the groups are formed through the process of random allocation, this will eliminate bias (except chance bias) which makes it possible to make inferences about an intervention which are simply not possible using other research methods (Torgerson and Torgerson, 2007).

The research discussed in Chapter 2 - *Computers in the Classroom* revealed that the quality of educational effectiveness trials has been the subject of much criticism in the area of ICT in Literacy teaching. Recently, there has been a greater international awareness among policy makers funders and practitioners of the need for researchers to establish, in an unbiased way, whether or not educational interventions (teaching programmes and practices, strategies and methods) are actually effective in improving children's educational outcomes (Torgerson, 2009). Based on the rigour, reliability and repeatability of the RCT approach, this is the method of choice for the evaluation of the educational effectiveness of the SPEL application. However, there are many biases to consider and limit; if these variables are uncontrolled they will significantly affect the outcome of the experiment. Biases reported by (Torgerson and Torgerson, 2003, Torgerson and Torgerson, 2007, Torgerson, 2009, Torgerson and Zhu, 2003, Brooks et al., 2006) are summarised here:

4.2.1 **Chance bias**

Randomisation will ensure that two or more groups are similar in most respects except by chance. The problem of chance bias can be minimised by stratifying, or "matching" participants on major predictive factors such that the subsequent randomisation will be balanced (Torgerson and Torgerson, 2003). Two such factors in educational experiments are gender and age; the management of gender and age bias is covered in Section 4.2.2 and Section 4.2.3.

4.2.2 **Gender bias**

There may be a difference in performance between boys and girls (Section 2.6) so an imbalance in gender distribution could bias the result. In educational trials carried out in mixed-sex schools, gender bias needs to be controlled. This can be done by "matching" children in terms of gender; forming intervention and control groups through random allocation will then ensure the groups are balanced in respect of this characteristic at the start of the experiment (Torgerson, 2009).

4.2.3 **Age bias**

As there can be up to a year's difference between the youngest and oldest child, it is possible that the older children will perform better than the younger children because they are more intellectually developed or have a greater relative level of experience. Age bias can be controlled by “matching” children of similar ages; forming intervention and control groups through random allocation will then ensure the groups are balanced in respect of this characteristic at the start of the experiment (Torgerson, 2009).

4.2.4 **Selection / Familiarity bias**

Randomisation eliminates selection bias which may otherwise occur if groups contain children that have fundamentally different characteristics that could explain differences in outcome (Torgerson and Torgerson, 2003). Selection bias may occur in educational experiments where people override the random allocation for reasons of personal preference. If a person familiar with the children, such as their teacher, selects the groups, he or she may, even unconsciously, introduce bias. For example, perhaps the knowledge that certain children are known not to work well together or others have a particular desire to work with a computer or any number of apparently minor attributes could influence the selection. Selection bias can be eliminated by the randomisation being carried out by a person who has no knowledge of the participants. Selection bias can be avoided by using a third party who has no vested interest (intellectual or financial) to perform the random allocation procedure (Torgerson, 2009).

Randomisation eliminates selection bias as long as all the participants are retained within their randomised groups for the duration of the experiment and during analysis of the results. People may wish to change groups following allocation for ethical, educational or administrative reasons. However, in order to preserve the original randomisation, the analysis must be carried out on the original groups. This is known as ‘intention-to-treat’ (ITT) analysis (Torgerson, 2009).

4.2.5 Contamination bias

Craven et al (2001) call this a "diffusion" effect and Brooks et al (2006) refer to it as "contamination". It is where one or more group members learn something about what the other group is doing which then affects the way they perform. Contamination bias can be avoided in computer experiments by preventing access by the control group to the computer application. This can be achieved in a number of ways. For example, the application can be put on laptop computers which are removed from the school between experimental sessions, or teaching staff can be asked to sign an agreement stating that the control group will not be allowed access to the computer application until the experiment is finished or perhaps desktop icons can be avoided to prevent easy access to the application.

4.2.6 Performance bias

Performance bias can occur if participants or teachers seek alternative treatments for the control group following the randomisation process (Torgerson and Torgerson, 2003). Parents or teachers may seek extra tuition for a child if they feel that the child is being disadvantaged by not being offered the intervention. Providing a "waiting list" approach is suggested by (Torgerson and Torgerson, 2007) to reduce this effect; this is where the control group are given the opportunity of using the computer program once the trial period has ended.

4.2.7 Dilution bias

Dilution bias can occur if the intervention treatment is not delivered adequately which may skew the experimental results by affecting the performance of the intervention group in a negative way (Torgerson and Torgerson, 2003). In educational experiments comparing the effectiveness of a computer intervention to classroom taught sessions, this type of problem can be avoided by developing computer applications which are usable with little or no instruction from members of staff running the laboratory sessions.

Conversely, dilution can occur if the control treatment is not delivered adequately. In experimental trials comparing the effects of a computerised intervention to a teacher-led session, this bias can be minimised by asking the teacher to deliver the same content in each experimental session as that being provided by the computer application.

4.2.8 Demoralisation effect

Some children may have become excited at the thought of using a computer and a new program only to find that they have been allocated to the control group. This may demotivate them and affect the results of the control group in a negative way. Providing a "waiting list" approach is suggested by (Torgerson and Torgerson, 2007) to reduce this effect. The control group can use the computer program once the experiment has been completed.

4.2.9 Reporting bias

Reporting bias occurs when researchers are more assiduous in their reporting of events in one group compared to another. This may happen for example, if a researcher wishes to report a positive outcome to their own experiment. This can be avoided by ensuring an external check of the data is made before the analysis is carried out. Reporting bias may also occur if a data collector consciously or unconsciously reports a biased outcome. This can be minimised if reporting is undertaken blind to treatment allocation; in other words, the person measuring the outcomes does not know which group the child is from (Torgerson and Torgerson, 2003).

4.2.10 Attrition bias

Attrition is the term given to participants who are lost prior to follow up measures.

Any amount of attrition can lead to selection bias unless attrition is a random event; participants may leave one arm of the study due to preference for example. In school trials, it is therefore important that participants are encouraged to remain in the original randomised groups for the duration of the experiment. A 'waiting list' design could help retain children in the control group if the reason for attrition was demoralisation.

Attrition bias can be minimised if assiduous follow-up is carried out, such as encouraging attendance of randomised participants at post-tests to ensure as few participants as possible are lost to follow-up (Torgerson and Torgerson, 2003). It is also highly likely that some children will be absent from school during the post-test data collection exercise. To avoid attrition bias caused by missing post-test results, it is important that the data collectors return to school to carry out the missing post-tests.

Attrition is almost inevitable in a substantial trial yet a literature search shows no common consensus on how to deal with it; (Leon et al., 2006) concur with this finding. There are various approaches to attrition: ignore it or don't report it; remove the participants from the analysis; carry out the analysis twice using extreme values for the missing scores or estimate a score. A popular or even recommended approach is to use Intention To Treat (ITT) analysis. However, the Consolidated Standards for Reporting Trials (CONSORT) RCT guidelines are currently being updated to remove the term "Intention to Treat" citing it as "*a widely misused term*". That is not to say that they do not approve of the ITT method but are aware of its misinterpretation (Schulz et al., 2010). Polit and Gillespie (2010) recommend the use of ITT analysis but also acknowledge that there is still some confusion among researchers as to what it really means. They attempt to clear this up with a detailed discussion of what they consider to be classic (true) ITT by stating: "*A true or classic ITT is one that removes none of the subjects from the final analysis …*".

Although Polit and Gillespie accept that random attrition will have minimal effect, it is very difficult to prove that the missing data affects no variables. They recommend the use of statistical methods to effectively create likely data using software such as the "Missing Values Analysis (MVA) module" available for SPSS. However, their main recommendation is to design the trial to minimise attrition and this is also the recommendation of Hutchison and Styles (2010) in their guide to running RCTs in educational research.

The possibility of attrition bias due to random events can be minimised (but not eliminated) if attrition rates are similar between the arms of a trial. Torgerson (2007) suggests as a rule of thumb, less than 5% attrition is not a problem. Fewtrell et al (2008) also suggest that attrition of $\leq$ 5% is usually of little concern. This is particularly true if the attrition is shared across the arms of the trial. However, some trials have to deal with high levels of attrition (> 20%) so MVA may be appropriate in those cases.

There are constant reminders in the literature that significant attrition could reduce the statistical power to an unacceptably low value (Leon et al., 2006, Polit and Gillespie, 2010, del Boca and Darkes, 2007, Fewtrell et al., 2008), so it should be recalculated to ensure that is not the case

4.2.11 **Bias control in the SPEL RCT**

The approach used to manage each bias in the RCT is detailed in Chapter 6 - *SPEL Usability and Phonemic Awareness Experiments*. After minimising bias as far as practicable, the groups for the randomised controlled trial were allocated using random sampling stratified on age and gender.

4.3 Qualitative evaluation methods

This section documents the qualitative methods used in the experiment to evaluate the usability of the computer application developed for this project; details of the application of these methods in the SPEL usability evaluation can be found in Chapter 6.

In 1997, Hanna, Risden and Alexander published work on usability testing of computer applications with young children, noting that:

> *"Traditional measures of usability such as productivity indices and speed and efficiency of task completion are not generally appropriate to use for children's products."*

(Hanna et al., 1997)

Hanna et al (1997) reported from their usability evaluations that young children could concentrate for approximately 30 minutes, so the activities in this study were designed to take approximately 20 minutes which allowed for children who liked to take their time or experiment with the interface. If a child was required to participate in longer sessions, a reasonable break would be given, or morning and afternoon sessions would be scheduled if it was not important that the child needed to remember the early work in the later session.

Jensen and Skov (2005) note from their evaluation of the literature relating to children's technology that research in this area has a strong focus on natural setting environments and is therefore typically conducted in schools primarily for evaluating educational products. The strong focus on field studies in a real world context is useful and necessary to understand the usability of products in the environment in which they will ultimately be used. In addition, when evaluating children's technologies, the most obvious way to recruit subjects is to place the evaluation in a school environment. The subjects for the evaluation sessions in this study were recruited from primary schools and the usability evaluation sessions took place in primary school computer suites.

A variety of usability evaluation methods were reviewed and are discussed in the following sections with regard to their suitability and usefulness to this project. The appropriateness of qualitative evaluation methods for collecting meaningful data from young children has been demonstrated in large early-literacy projects. Nutbrown and Hannon (1993) argue that new measures to assess children's early literacy should include techniques such as interviews and questionnaires; these techniques were used successfully to collect meaningful data for the Raising Achievement in Early Literacy (REAL) project (Nutbrown et al., 2005). Similarly, findings from the Peers Early Education Partnership (PEEP), which investigated the effects of a literacy program on the children and families from the community in which it was implemented were gathered by the same qualitative research methods (Evangelou et al., 2005). Nielsen et al (2002) recommend data collection from more than one source but point out that the time to analyse and the cost of collection should be considered; more than one method was used in the evaluation of SPEL to improve the reliability of the results. Potentially relevant evaluation methods are discussed in the subsequent paragraphs and include the rationale for using or not using a particular method.

4.3.1 Observation

During the evaluation of applications developed for this project, the child-evaluators' interactions were observed. It was noted during early evaluation sessions that children tended to "play up" to a video camera placed in front of them, so a video camera was placed behind the child to view the screen and the way that the child interacted with the computer. This worked well, as the children tended to forget that the camera was there. To support the video recordings, note taking was carried out as part of the procedure.

4.3.2 Action tracking

An action tracking facility captures detailed interaction data to back-up the visual and audible observation data captured. This feature enables the evaluator to replay the key strokes made by the child; it can "play back" the child's interactions by recalling all the mouse movement, mouse clicks and timings. This data is fed back into the application to replicate the child's interaction. This feature can be useful for example, to look at a particular usability issue which may have been obscured from the camera's view or was missed by the observer. An action tracker was developed and incorporated into the application as part of this project to assist with the interaction evaluation phase; details can be found in Section 5.4.1- *SPEL High level technical details* and Appendix G - *Capturing interaction and performance* data.

4.3.3 Thinking aloud, Constructive Interaction and Co-discovery

In usability evaluations, verbal protocols are claimed to uncover the cognitive processes of test participants (Edwards and Benedyk, 2007). The think-aloud method (Markopoulos et al., 2008) involves the user vocalising their thought processes during the interaction; adult users normally require occasional prompting. Research carried out by Donker and Reitsma (2004) has shown that children who think-aloud during testing uncover more problems than children who answer specific questions. However, the same study reported that children frequently need to be prompted to keep talking. As prompting may cause children to feel obliged to mention problems to please the experimenter, this could lead to non-problems being reported. When children are instructed, but not prompted to talk or think aloud, they may be more comfortable and therefore report fewer non-problems. However, children may also fail to mention certain problems that they do not consider very important. The researchers suggest observing the behaviour of children while they are performing this kind of test.

Nielsen (1993) recommends that evaluators use a variation of think-aloud called constructive interaction which involves children working in pairs, as children may find it difficult to follow the instructions for a standard think aloud test. Constructive interaction, (Miyake, 1986), also known as co-discovery learning, (Kennedy, 1989) involves two test subjects who work together to try to solve tasks. Als et al reported on an experiment that compared think-aloud and constructive interaction in usability testing, that:

> *"Constructive interaction with pairs of children knowing each other identified more problems (on all severities) and specifically more critical problems."*
>
> (Als et al., 2005)

However, the children in their study were 13-14 years old and the researchers also reported that the children had no major problems in following the standard think-aloud protocol.

Constructive interaction was not used in the evaluation of SPEL, as early trials of this method showed that children as young as 5 years did not work well in pairs; either one child dominated the session, or the pair spent most of their time fighting for the mouse and pushing each other in order to get to the screen. Van Kesteren et al (2003) also found that young children tended not to cooperate well when using the co-discovery method.

The think-aloud method was successful in some aspects of data collection during the usability evaluation of SPEL, but not in others. For example, children tended to think aloud until they became engrossed in the task at hand, then they became silent until prompted; this may have been due to the increase in cognitive load. Therefore for the evaluation of the SPEL interface, when children were not struggling the evaluator simply asked the child to think aloud, whereas if a child was obviously struggling or deeply engrossed, focused questions such as “Which bit are you finding difficult?” were asked.

4.3.4 **Performance measures**

Quantitative measures tend to be task-based. This approach was not used with the children in the interface evaluation phase of this project because the evaluation sessions were designed to make the children feel at ease in the knowledge that they were not being tested; if they are given tasks to do which are designed to test the application and the application doesn’t work properly, the child is likely to assume that they did something wrong which may make them feel uneasy about continuing. However, adults carried out task-based evaluation of supervisor screens (screens inaccessible to the children for administration and preference settings) but these evaluations have not been reported in this thesis as they are not directly relevant to the Child-Computer Interaction.

4.3.5 **Retrospection and post-session interviews**

Retrospection requires that participants comment on their thought processes after tasks have been completed. The main problem with carrying out this procedure directly at the end of a session is that children who are tired will be less likely to provide useful feedback. The problem with carrying out this procedure later in the day, or on another day is that the children are less likely to be able to remember much of the detail of the sessions and therefore less likely to be able to provide useful feedback. Although Van Kesteren et al (2003) suggested that a good way of prompting the memory of the user is to review the video tape of the session with the participant, this method was not considered practicable with young children and therefore not used during the evaluation of SPEL’s interface. The videos were however reviewed by the evaluator and post-session data was collected through questionnaires and interviews.

4.3.6 Post-Evaluation Questionnaires and Interviews

Questionnaires for young children need to be designed differently than questionnaires for adults; it is unreasonable to present questions in text form to beginning-readers and according to Read and MacFarlane (2006):

> *"asking good questions is not easy, and for some children, understanding and interpreting the question, and formulating an appropriate response can be very difficult."*

Read and MacFarlane (2006)

These researchers recommend the use of "The Smileometer", a discrete Likert type scale illustrated in Figure 4-1.

Figure 4-1: "Smileometer" questionnaire gauge

The Smileometer was used to elicit questionnaire feedback from the children during interviews in the SPEL evaluation sessions as discussed in Section 6.2.3 - *Evaluation feedback interviews*.

4.4 Research ethical considerations

Fundamental ethical considerations are: safety, consent, honesty and privacy.

Ethical approval: Full ethical approval was obtained from the University of Central Lancashire research office, which documented that all measures were in place to ensure the safety of the children involved in the experiment.

Safety: Technical equipment such as such as lap-top computers and video cameras can be dangerous with young children, who can be inquisitive and lively. Safety in the school is discussed initially with the headteacher and then with the class teacher. The position of equipment and cabling was made secure and safe before the usability experimental sessions involving the children began. For personal protection of a child and the researcher, approval from the Criminal Records Bureau (CRB) was obtained for the researcher and data collectors working with the children.

Consent: Parental consent for children involved in the experiment was gained by the head teacher of each of the schools involved.

Privacy: To ensure no personal data was held on the children, a list of ID numbers was obtained from each school office to avoid recording the names of the children.

Deception: In the context of design and evaluation of computer applications, deception is used to describe a situation where the facilitator is untruthful to the evaluator. One of the most typical types of deception using computers is a Wizard-of-Oz scenario where, unknown to the computer user, the prototype computer application has a hidden human operator (Leventhal and Barnes, 2007); there was no such deception used in this project. However, a form of deception was used to avoid biased responses from the children during evaluation sessions and in post-session interview results; children were asked to evaluate the computer application for the University and not for the researcher. The researcher distanced herself from the application development to avoid the child feeling obliged to give positive responses. According to Markopoulos et al (2008) this can be classed as good practice rather than dishonesty.

4.5 Conclusions

Field studies involving young children need to be planned carefully. Safety and ethical considerations are paramount when working with this lively and inquisitive age group. The research and evaluation methods documented in this chapter are put into practical use in Chapter 6 - *SPEL Usability and Phonemic Awareness Experiments* which documents the use of the qualitative and quantitative methods discussed in this chapter. However, before reporting the evaluation of SPEL's usability and educational effectiveness, the development and operation of this phonemic awareness application is discussed in the following chapter: *SPEL Operation and Interface Design.*

Chapter 5

SPEL Operation and Interface Design

Chapter 5 SPEL Operation and Interface Design

5.1 Introduction

The shortage of randomised control trials to evaluate the effectiveness of computerised phonics programs as learning tools in UK primary classrooms was identified in Chapter 2 - *Computers in the Classroom*, which also discusses the shortage of usability evaluations of computer software used in computer effectiveness trials and the importance of this aspect of software development. Usability evaluations of existing phonemic awareness computer software packages could have been carried out as part of this project, but this would have been a time consuming process and could have involved setting up usability evaluation trials for a number of applications with no guarantee that a suitably effective one would be found. Therefore, the decision was made to develop a bespoke application which would be subjected to usability testing and improved through iterative refinement. SPEL is a prototype computerised tutor which has been developed to teach phonemic awareness through segmenting and sequencing activities.

The usability effectiveness of the SPEL application is reported in Section 6.2 - *SPEL Usability experiment.* Findings from the evaluation were used to refine the application before using it in a randomised controlled trial designed to evaluate the educational effectiveness of the application as a phonemic awareness tool.

The interface to the application was designed using the child-computer interface design guidelines developed as part of this project and detailed in Appendix A - *Child-Computer Interaction*. Reference to the design guidelines and how they have been used to inform the design of SPEL's interface is detailed in this chapter.

5.2 Generating natural human speech output

The age range of children targeted in this project is 5 to 6 years. As the children are either non-readers or beginner-readers the interface cannot rely on text as an output mechanism for instructions, feedback or help. Research detailed in Appendix A - *Child-Computer Interaction,* suggests that spoken output is an effective mode for communicating instructions, help and feedback to young children. There are two modes for spoken output: speech synthesis and recorded human speech. Teachers consulted in this project preferred human speech for young users as it is clearer and uses the correct pronunciation and intonation. Therefore, human speech has been used for the SPEL application.

The work of Balogh (2001) shows that natural prosodic units of speech are preserved when using human speech in computer interfaces. These findings are of particular relevance to a system that relies heavily on the correct pronunciation of phonemes and words. Incorrect or inconsistent pronunciation of phonemes by teachers has been reported as a problem which can lead to the confusion and frustration of young children learning the alphabetic code; Rose (2006) reported that imprecise pronunciation of phonemes by adults was found to be a problem area in the teaching of phonics in some of the schools reviewed by the HMI. The use of human speech in the SPEL application can provide consistent and correct pronunciation of phonemes and whole words.

Kehoe and Pitt (2006) point out that "*Care should be taken in selecting a voice so that the speech persona is consistent with the application*". The narrator for the instructions, help and feedback utterances for SPEL is a young child. The prosody of speech for these utterances was preserved by creating a script for the narrator; she was asked to speak the highlighted words in an "upbeat" manner. Table 5-1 provides an excerpt from such a script which was developed during discussion sessions with teachers to provide a dialog that was as natural and as close as possible to the customary teacher-child dialog used in the classroom teaching phonics.

Required phrases with word to speak highlighted	**Spoken word or phrase**
"You're <u>fantastic</u> at this."	"fantastic"
"You're <u>brilliant</u> at this."	"brilliant"
"You're <u>great</u> at this."	"great"

Table 5-1: Application speech script example

By recording single words or very short phrases in this way, the key word can be selected randomly by the application and concatenated into the rest of the phrase. After recording the common parts of the phrase once (i.e., “You’re” and “at this”), the child would read the common element of the phrase mentally but actually speak the blue part with the intonation it would have if spoken within the whole sentence. The word “fantastic”, “brilliant”, “great” or other appropriate word is then selected by the application and concatenated onto the phrase “You’re”. It then concatenates “at this” to make up the whole sentence. This enables variable feedback phrases but avoids excessive recording time and computer storage space.

Davison et al (2005) confirm that: “*the cost in time and money to create and update human speech recordings can be very significant*”. Careful scripting and concatenation of phrases can significantly reduce development time and still produce a variety of phrases which sound natural when played by the application.

This type of careful scripting can also reduce the workload on the narrator, which is of particular importance if the narrator is a child. For example, careful scripting for the three phrases in Table 5-1 would mean that the narrator is able to record the “upbeat” words in one session. In this particular example, the child would be required to speak five short phrases instead of three long ones; the obvious benefit to this method is that in recording say ten long phrases, the child would still only be required to speak twelve short phrases. Niemi and Ovaska (2007) reported when working with children recording natural voice for computer interfaces that the children were able to work for about an hour without becoming bored; hearing their voices become part of an interface that would benefit others seemed to motivate them. However, Hanna et al (1997), recommend that half an hour is the maximum length that children can be expected to work on design or evaluation activities. During voice recording sessions for applications developed as part of this project, it was found that the attention span of the child narrators was about half an hour.

5.3 Implementation of child computer interaction guidelines

This section introduces the interface design guidelines developed as part of this project; development of the guidelines is detailed in Section A2 - *Design Guidelines*

Development. The design guidelines were followed when designing the interface for SPEL. Each guideline is italicised then followed by a brief explanation of its application to SPEL:

Minimise the use of text as a feedback mechanism. All feedback is spoken or graphical.

Minimise the number of interactive controls. There is no caption bar or control icons for minimise, maximise and close, so the child cannot inadvertently put the application into a mode which could confuse them. The only interactive controls used in SPEL are the graphemes on the screen and the characters: “Floppy dog" and “Ducky”.

Vary the feedback. To provide natural dialog and interesting feedback, all non-graphical feedback consists of concatenated human speech made up from words and phrase combinations chosen randomly at run time from a carefully prepared list to produce appropriate feedback sentences.

Minimise keyboard input. No keyboard input is required; the software is aimed at 5 to 6 year old children, so the interaction between child and computer is limited to mouse clicks and rollovers.

Make use of randomness. Feedback phrases are randomised to maintain interest and provide a more natural Child-Computer Interaction experience.

Enable early exit from the application but make this inaccessible to the children. Children may exit the application early either accidentally or intentionally. To avoid this happening, the application may be terminated early using a key combination known only to the teacher.

Provide tiered pro-active interactive help. The SPEL application originally used three tiers of help:

1. An interactive help character ("Floppy dog") was provided on the screen for the children to request help (discussed in Section 5.4.2).
2. If the child made an incorrect choice, the application would output an appropriate response such as "try again", or "try another".
3. The application intervened after a pre-set number of errors such that the application "took over" and said to the child "let me show you this one" - the application showed the child how to solve the problem then "remembered" to ask the same question again at a later time.

However, a decision was made to leave only the on-screen help character in the application following usability evaluation sessions documented in Section 6.2.

Enable user defined options but make them inaccessible to the children. This guideline is concerned with ensuring that applications are as flexible as they need to be. The teachers involved in this project stated that a feature to add new words into the application would be very useful but it needed to be easy to do. The feature to add new words by saving and loading new text files was implemented in the prototype to facilitate changes to the words used by the application without modification and recompilation of the program. Detail of this flexibility provided by SPEL is provided in Appendix E - *SPEL Application Configuration*.

Ensure that interaction hardware is of relevant size for the user

A small infra-red mouse was provided as this type of mouse is easy to operate with small hands and does not need a mouse mat which enables the child to utilise any free desk space.

Design for no scrolling

This guideline was developed to ensure that the whole screen was displayed at all times and the screens displayed nothing but relevant easily accessible content. There are no menus or scroll bars in SPEL.

5.4 SPEL Architectural overview

SPEL is a prototype application designed using the experience gained through researching phonics and experience gained from studying interface design for young children. Research carried out into the principles and practice of high-quality phonics teaching discussed in Chapter 3 - *Phonics in UK classrooms: the debate*, has been used to inform the scope and delivery of content of the SPEL application. The application provides a systematic and consistent platform. There is no time limit on activities so that children can experiment with the sound↔symbol correspondences and work at their own pace. The phonemes, graphemes and words used by the application are based on the DfES Letters and Sounds phonics scheme (DfES, 2007b) and are documented in Appendix B. Following the year 1 teaching plan for the Letters and Sounds scheme provided the opportunity for SPEL to progress from teaching simple well known words onto less familiar more complex words formed in the first-middle-last order.

5.4.1 **SPEL High level technical details**

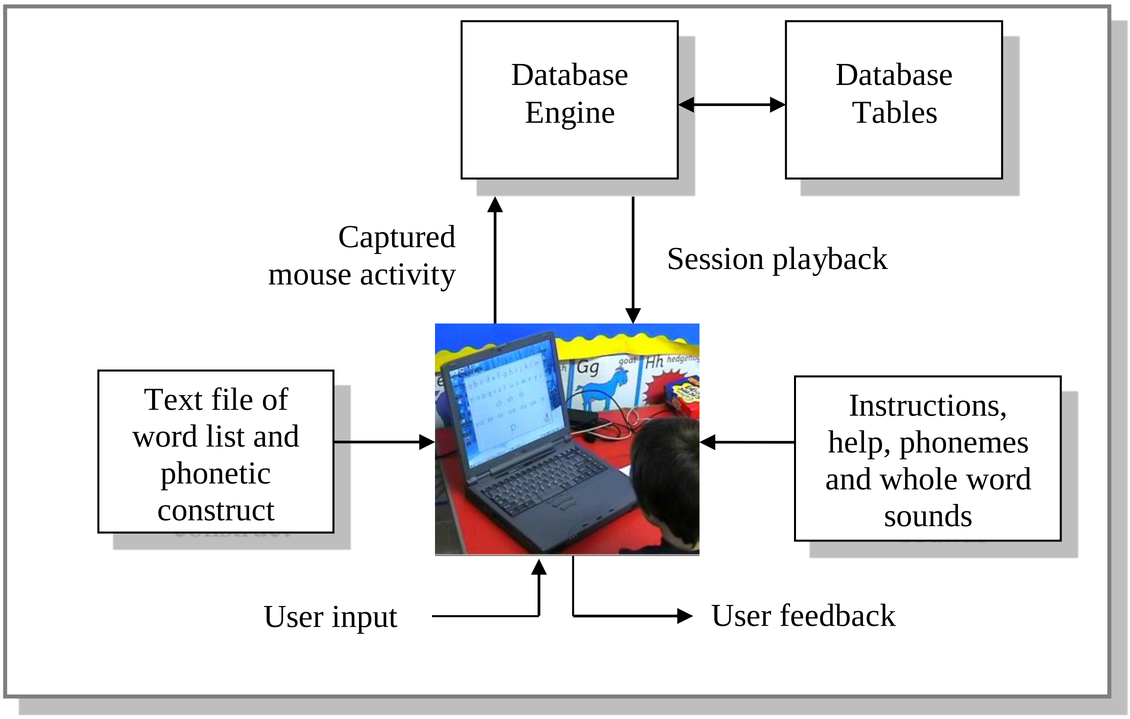

Figure 5-1: SPEL application block diagram

The author developed the application and linked it to a Paradox database engine. It loads a list of words and their phonetic elements from a text file. The contents of this file are used to display or utter individual words to be sequenced or segmented (detailed in Appendix E - *SPEL Application Configuration*). The vocal feedback, including reference utterances of each of the words and phonemes, is loaded into the application; a recording facility to enable teachers to add their own words would be integrated into a finished application. The database engine writes to tables to record each mouse movement and each mouse click made by each child on a per-session basis; this data is then available for analysis. Further usability analysis is enabled by the development of another hidden application feature that can take the recorded keystrokes, mouse clicks and timing information to replay the child's session; this can be useful to observe usability issues which were not necessarily captured clearly on camera. An example of the data captured in one of the tables is illustrated in Figure 5-2 and a description of the data capture process and table columns is provided in Appendix G - *Capturing interaction and performance data.*

Table : D:\Work (Z)\...\Capture.DB

Capture	IdNumber	SessionNum	Object	DownTime	DownX	DownY	DownString	UpTime	UpX	UpY	UpString	Button
1	2	1	lbC	14:28:28	15	51	c	14:28:29	15	52	c	Left
2	2	1	lbA	14:28:45	11	50	a	14:28:46	11	50	a	Left
3	2	1	lbT	14:28:58	10	44	t	14:28:58	10	44	t	Left
4	2	1	lnV	14:29:12	4	44	v	14:29:12	4	46	v	Left
5	2	1	lbA	14:29:14	10	46	a	14:29:14	10	78	a	Left
6	2	1	lbN	14:29:16	14	40	n	14:29:16	16	40	n	Left
7	2	1	lbB	14:29:27	6	40	b	14:29:27	7	40	b	Left
8	2	1	lbB	14:29:29	7	40	b	14:29:29	7	40	b	Left
9	2	1	lbI	14:29:32	8	36	i	14:29:33	-16	50	i	Left
10	2	1	lbG	14:29:34	14	50	g	14:29:34	14	50	g	Left
11	2	1	lbH	14:29:47	15	60	h	14:29:47	11	40	h	Left
12	2	1	lbO	14:29:49	11	55	o	14:29:49	11	55	o	Left
13	2	1	lbT	14:29:51	7	43	t	14:29:51	8	38	t	Left
14	2	1	lbH	14:30:04	23	59	h	14:30:04	22	59	h	Left
15	2	1	lbA	14:30:05	16	47	a	14:30:05	58	71	a	Left
16	2	1	lbT	14:30:07	14	53	t	14:30:07	15	53	t	Left
17	2	1	lbL	14:30:19	6	33	l	14:30:19	6	33	l	Left
18	2	1	lbI	14:30:26	9	37	i	14:30:26	8	38	i	Left
19	2	1	lbP	14:30:27	18	46	p	14:30:27	18	46	p	Left
20	2	1	lbB	14:30:34	3	53	b	14:30:34	5	53	b	Left
21	2	1	lbA	14:30:35	22	50	a	14:30:35	22	50	a	Left
22	2	1	lbG	14:30:37	7	64	g	14:30:37	7	64	g	Left
23	2	1	lbS	14:30:47	16	48	s	14:30:47	16	48	s	Left
24	2	1	Image1	14:30:48	496	72	Background	14:30:49	496	78	Background	Left
25	2	1	lbI	14:30:53	5	42	i	14:30:53	-173	194	i	Left
26	2	1	lbT	14:30:55	8	33	t	14:30:55	8	33	t	Left
27	2	1	lbS	14:31:04	16	54	s	14:31:04	26	54	s	Left
28	2	1	lbI	14:31:07	3	44	i	14:31:07	5	45	i	Left
29	2	1	lbX	14:31:09	10	38	x	14:31:09	7	40	x	Left
30	2	1	lbH	14:31:18	2	42	h	14:31:19	0	42	h	Left
31	2	1	lbO	14:31:20	10	33	o	14:31:20	35	34	o	Left
32	2	1	lbT	14:31:21	5	38	t	14:31:21	5	38	t	Left

Record 30 of 96

Figure 5-2: Example database table created by the Sequencing activity

5.4.2 **SPEL Sequencing and Segmenting Activities**

SPEL offers segmenting and sequencing activities, based on Brooks' *"Conceptual schema of synthetic phonics"* (Brooks, 2002a) illustrated in Figure 5-3.

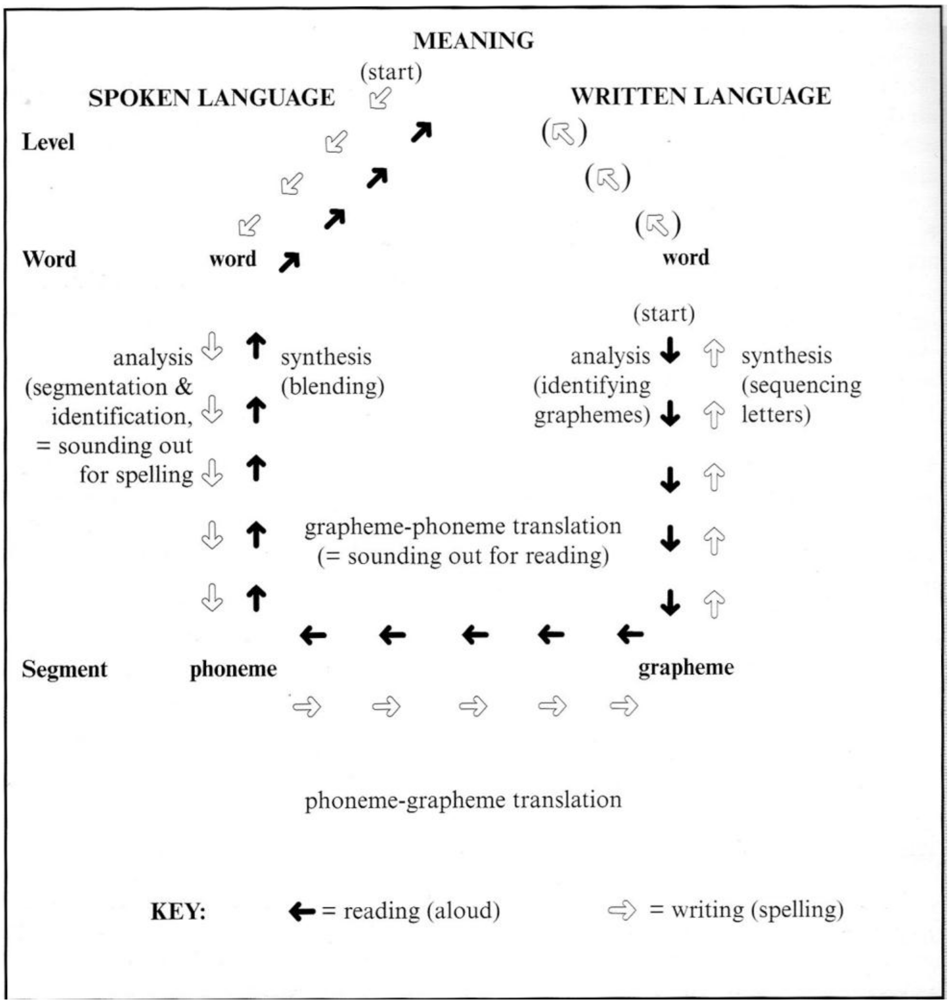

Figure 5-3: Conceptual schema of synthetic phonics

(Brooks, 2002a)

SPEL's sequencing (word building) activity requires the child to listen to a word, mentally segment the word into its constituent phonemes, identify the corresponding graphemes then click the graphemes in the correct order to sequence the word on the screen.

Each time a grapheme is clicked, it is sounded-out by the computer; if this produces the correct phoneme, the grapheme is added to the word until the complete word is built up and spoken on completion by the computer. In practice all the stages may overlap and be achieved so rapidly that subjectively they seem to take place simultaneously. The child is able to explore the sound↔symbol mapping by trial and error if necessary; each time an incorrect grapheme is clicked, because it is sounded out, they should eventually become familiar with the sound↔symbol mapping and observations in trials have shown that they quickly become more selective in their choice - even the slower readers quickly converged onto the correct phoneme selection as they built up a mental database of the sounds and symbols.

An example of the sequencing screen is illustrated in Figure 5-4 - Sequencing the word “like”. If the child clicks on the “Ducky” help icon, they will hear the word again. If they click on the “Floppy Dog” icon, the application will highlight and sound-out the correct graphemes before encouraging them to try again with the same word. If the word contains a split-vowel digraph, the concept is explained then an example word is shown being sounded-out and built up before encouraging the child to finish off the current word.

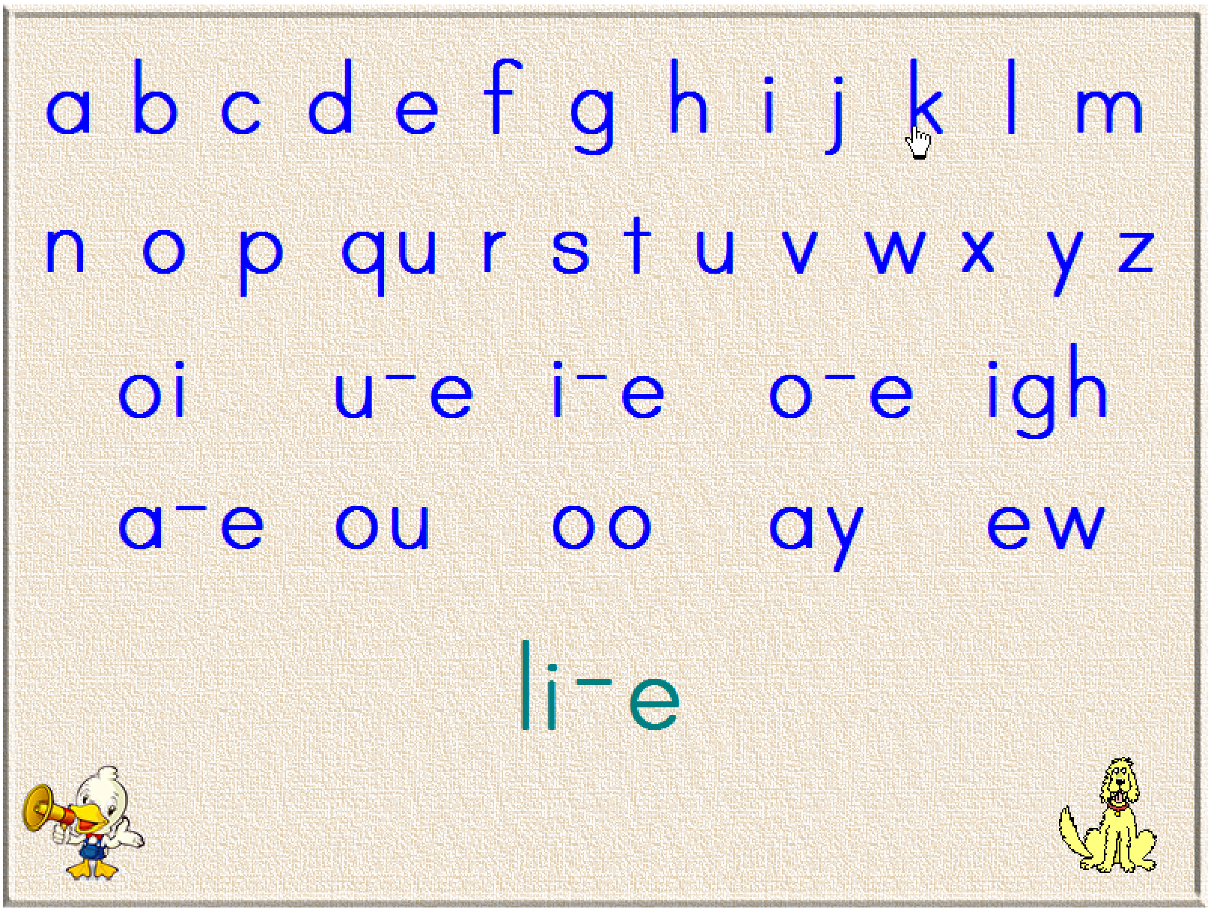

Figure 5-4: Sequencing the word "like"

Figure 5-4 illustrates the correct selection of the grapheme <l> and split vowel digraph <i-e> and the cursor positioned to select the final grapheme "k" in response to sounding out the word "like".

The second activity available in SPEL teaches segmenting (decoding of words). A word is displayed. A child clicks on the graphemes which make up the word in sequence. The child needs to mentally segment the word to identify the individual graphemes. An example of the process is shown in Figure 5-5 which illustrates the word "saw" being segmented.

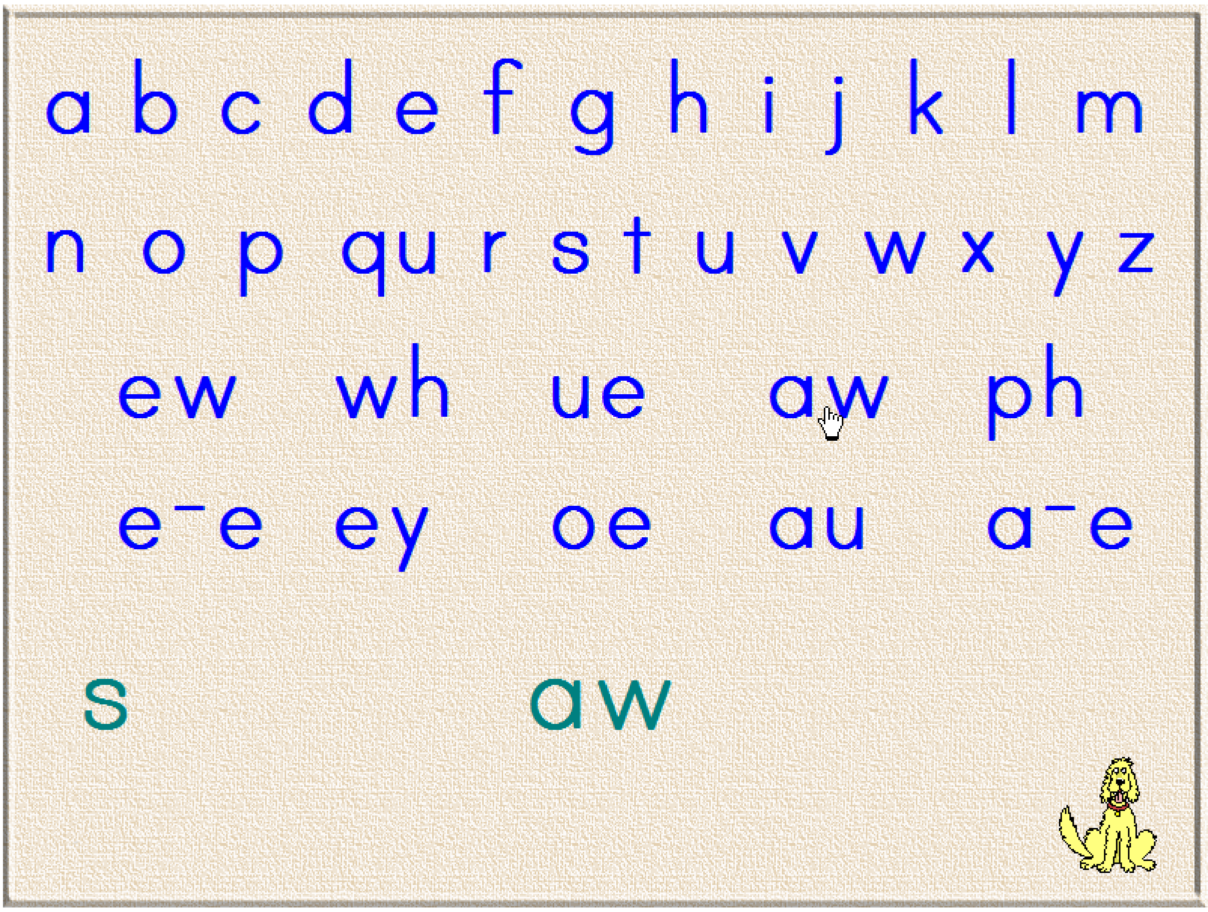

Figure 5-5: Segmenting activity breaking down the word "saw"

The word is displayed but not sounded. The child needs to mentally break down the word into its phonemes, map the phonemes to graphemes then click the graphemes to decompose the word. When a grapheme is clicked it is sounded out whether it is correct or not; this forms part of the sound↔symbol mapping learning process. If the grapheme is correct, it is split from the word and moved to the left. The example shows the word "saw" partially segmented; the child would be required to click on the final grapheme <aw> to complete the word. On completion, the word is sounded to confirm the mapping from the written word to the spoken word. If the child needs help, clicking the "Floppy Dog" help icon will cause the application to highlight the appropriate graphemes before encouraging the child to complete the word; the word is not sounded in the help sequence but the graphemes are sounded and highlighted. The repeat word icon "Ducky" is not available in this activity as the word is not sounded until segmentation is complete.

With the application designed and implemented, experiments were set up to determine its usability and educational effectiveness. The implementation and interpretation of the experimental evaluations are the subjects of the next chapter: *SPEL Usability and Phonemic Awareness Experiments*.

Chapter 6

SPEL
Usability and Phonemic Awareness Experiments

Chapter 6 SPEL Usability and Phonemic Awareness Experiments

6.1 Introduction

This chapter documents three experiments involving young children aged 5 to 6 years from several primary schools: a qualitative interface evaluation experiment involved twenty children to determine the ease with which SPEL could be used by young children (Section 6.2); a quantitative pragmatic Randomised Controlled Trial (RCT) pilot which involved 19 children from a single school to test SPEL as a phonemic awareness tutor (Section 6.3.1) and a follow-up quantitative pragmatic RCT involving 266 children from four schools in the North West of England designed to more accurately determine the effectiveness of SPEL in developing phonemic awareness in young children (Section 6.3.2). Lessons learnt from the pilot RCT were reflected in the design and execution of the main RCT.

To ensure that the results of the RCT are as valid as can be reasonably achieved, as many potentially confounding variables as possible were identified and controlled; the variables and how they are dealt with are discussed within this chapter but one variable in particular, the usability of the SPEL interface, required a great deal of work to control it; this required a separate qualitative experiment to be undertaken involving twenty children and forms a contribution to this project and to the research field of Child-Computer Interaction (ChiCI).

If the interface to the application is not intuitive, the effort required by a child to use it would be likely to skew the results of the application's effectiveness – particularly with the less confident learners. However, providing an interface that looks intuitive to the developer is not enough to consider it neutral in the RCT experiment; it was necessary to research and develop the area of interface design for young children (discussed in Appendix A), and use the guidelines from Chapter 5 to develop the interface. The results of that work were used to design a qualitative usability experiment to ensure that the SPEL application was fully intuitive to children aged 5 to 6 years in order to discount usability as a confounding factor in the RCT.

The SPEL qualitative experiment is documented in the first part of this chapter as the results are used in the pragmatic RCT experiments documented in the second part of the chapter. The quantitative RCT experiments were designed to determine the educational effectiveness of SPEL as a phonemic awareness tool.

6.2 SPEL Usability experiment

The SPEL application was developed using the interface design guidelines discussed in Chapter 5 and Appendix A. It was tested using traditional software engineering methods then usability tested at a primary school in the North West of England. The usability process is summarised in Figure 6-1.

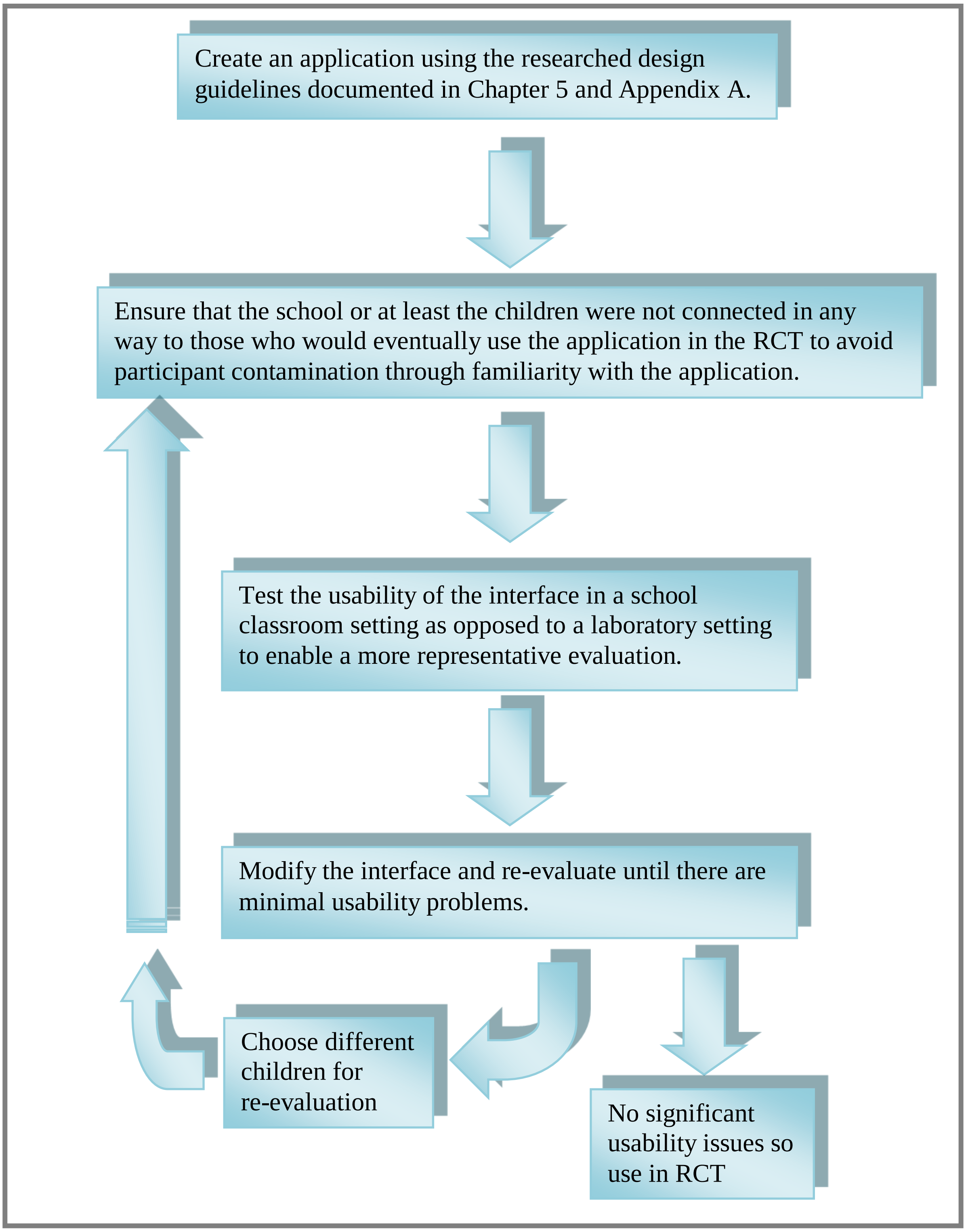

Figure 6-1: Interface usability evaluation lifecycle

As illustrated in Figure 6-1, each time the application is re-tested, a new set of children needs to be found. Fortunately SPEL only required two evaluation iterations, which may have been due to the large amount of research work directly leading to the interface design of the application.

The number of usability testers used in this evaluation was based on research findings on the subject. Faulkner (2003) reported that 10 usability testers uncovered most problems. However, Constantine (2003) and Faulkner (2003) both claim that the selection of users and rigour of the evaluation method is more important than the number of users. Donker and Reitsma (2004) recommend that gender issues are taken into account in computer evaluation studies. Based on these findings, 10 children were chosen to evaluate SPEL; 5 boys and 5 girls.

Each child was asked to segment and sequence the same set of words during the evaluation. Several methods discussed in this section were used in the evaluation sessions leading to the development of SPEL to improve the reliability of the qualitative analysis. Although Nielsen et al (2002) recommend that more than one usability data collection method is used, they also claim that the thinking-aloud technique is the most effective method. However, the results of the usability evaluation of SPEL did not demonstrate this, as the children found it difficult to use the application and think-aloud at the same time. The evaluator needed to continually prompt the child, which may have affected the way the child used the application. Therefore, a variety of evaluation methods were used such that they complemented each other to provide a greater level of assurance in the quality of data collected. The other methods that were found to be appropriate to this age group were: direct observation by a researcher and video recording reviews; action tracking where the child's mouse interaction was recorded by SPEL; post-session questionnaires using a simplified scoring approach which has been shown to be effective with young children.

The action tracking facility (Section 4.3.2) built into the software provided a useful backup for any data missing from the video recordings. Questionnaires (Appendix C) appropriate to this age group were designed and used to capture data during post-session interviews (Section 4.3.6); the Smileometer (Section 4.3.6) was found to be effective in capturing quantitative data from young beginner-readers. Co-discovery (Section 4.3.3) was tried but was found to be ineffective with this age of child; either one child dominated the session, or children fought for dominance of the session.

Think-aloud sessions (Section 4.3.3) were carried out with children using a lap-top computer to gain an insight into the child's thought processes with regard to navigation and use of the application. A small mouse was used as earlier application evaluation sessions used to develop the interface design guidelines (Appendix A) had shown that some children of this age found it difficult to use a regular sized mouse. An evaluation session with each child lasted about 15 minutes.

Following the methods discussed in Section 4.3.1, the sessions were recorded using a video camera placed behind the child; a script was used by the evaluator to ensure that each child received the same instructions on how to use the application. Observational notes were also taken by the evaluator. At the end of the session, the children were asked to complete the questionnaire in Appendix C; they did not need to write anything, they responded to questions by pointing at various positions on the "Smileometer"; the procedure is discussed in more detail in Section 6.2.3.

6.2.1 **SPEL usability issues from iteration 1**

The SPEL application originally used three tiers of help:

1. An interactive help character for the children to request help.
2. An incorrect choice would output appropriate spoken feedback.
3. After a pre-set number of errors the application demonstrated the process.

What actually happened during the evaluation session was a surprise; most children during the session were able to sequence and segment simple CVC words and for these children the help was not invoked. However, the few users who were less confident with the sound↔symbol correspondence and less confident in the skills of sequencing and segmenting clearly wanted to experiment by clicking on the graphemes to hear the sounds associated with them; one child even said he wanted to listen to the "talking letters". However, this invoked the feedback for each incorrect choice and invoked the pro-active help on making three incorrect choices. These children became frustrated and embarrassed by the constant feedback generated by the system when an incorrect choice was made which made them reluctant to explore.

On completion of the list of simple CVC words, such as <c> <a> <t> and CVC words containing consonant digraphs, such as <ch> <a> <t>, the system began to present words which required the child to sequence and segment CVC words containing vowel digraphs, such as <c> <oa> <t>. The majority of children who had easily completed the simple CVC combinations and with a little more effort the CVC combinations containing consonant digraphs, appeared at this point to be very confused. Most children immediately chose the single vowel grapheme that had the same sounding letter *name* as the digraph; for example, if they were required to click on the grapheme which corresponded to /əʊ/ sound in <c> <oa> <t>, they generally chose <o>, because they were familiar with the ‘owe’ sound of the alphabetic letter name of this vowel which is the phoneme /əʊ/. When the feedback from the system indicated that they had made an incorrect choice, they began to randomly click around the graphemes on the screen. Of course, the system began to generate the feedback on each incorrect choice and the pro-active help started on the third incorrect attempt. Because most of the children were not immediately able to make correct choices, the majority of children became frustrated and embarrassed by the constant feedback on their incorrect choice.

The issue was discussed at length with teachers and colleagues following the evaluation and two possible solutions were proposed: implement a facility to switch the choice feedback and pro-active help on or off or remove it from the application. It was decided that the choice feedback and pro-active help should be disabled until sufficient time could be allocated to properly research the area of interactive proactive help. A potential problem with disabling help is that the child could simply start at the first grapheme and click every other grapheme in turn until they heard the one they wanted. After discussion with the teachers it was decided that this may not be such a problem because the child would still be learning the sound↔symbol correspondences and that could only be a good thing.

An additional problem with help arose in early prototypes that used a lot of audio output for initial instruction on how to use the application. When it was evaluated, however, some children became frustrated by the long-winded instructions, with behaviour ranging from clicking around the screen in an attempt to get the tutoring session to start, to looking under the desk! Figure 6-2 shows a distracted child waiting for the instruction section to end. Based on these findings, the initial instruction was removed from the application.

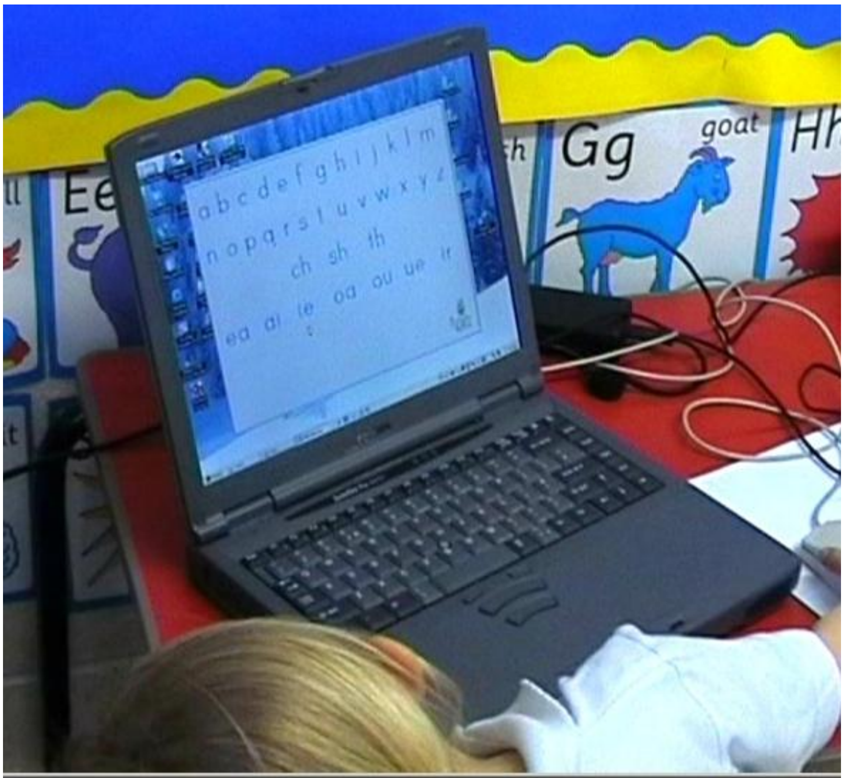

Figure 6-2: Child waiting for instructions to finish

Changes were implemented based on the in-class evaluation findings: the choice feedback, pro-active help and initial instructions were removed, but since the children liked the screen character "Floppy dog" it was left in the application. A second usability evaluation was then undertaken at the same school but used a selection of children not involved in the first evaluation.

6.2.2 **SPEL usability issues from iteration 2**

The removal of the initial instructions from the application meant that the children needed brief instructions (read from a script for consistency) on how to operate the application but other than that the same process was carried out.

No major usability issues were noted during this evaluation and all of the children completed the exercises without any usability problems. The decision to remove the choice-feedback, pro-active help and initial instructions improved the user experience in this particular application. This was validated by the children actively exploring the sounds by clicking graphemes. However, one child did forget what to do and without prompting he simply sat and looked at the screen. Prompting from a pro-active help engine should overcome this type of problem and it is suggested as an area for further work.

6.2.3 **Evaluation feedback interviews**

At the start of the evaluation sessions, it was explained to the children involved that they were not being tested, but that they were testing the system to see if they liked it. The researcher deliberately avoided associating herself with the application development to ensure that the children could feedback freely without fear of offending her. The children involved in the usability sessions of SPEL were asked a series of questions using the "Smileometer" chart to indicate their quantitative response; this is a useful tool when eliciting quantitative information from young children. The technique is discussed in Section 4.3.6 and the results are presented here.

The Smileometer was used to gather questionnaire feedback from the children in the evaluation sessions of this project so that they did not have to read any questions or write any answers. The expressions of the 5 smiley faces were explained to each child before the questions began. The same script of questions was used for each child to ensure consistency and to avoid recording bias by the researcher; each child applied a judgement score by pointing to the appropriate face on the Smileometer. The interviews took place immediately following the usability evaluation session to ensure that the children had not forgotten their experience.

The questionnaire was created by the researcher and checked by the supervisory team before being tested with school children; it was modified slightly before it was used with different children in the formal evaluation sessions. The only modification required was to change the question, “Do you have a computer at home?” into, “Do you have a computer at home that you are allowed to use?” as some children answered yes to this question but it was subsequently found that they were not allowed to actually use the computer.

Both the evaluation and questionnaire were facilitated by the researcher to maintain consistency for the children and to enable the researcher to gain an overview of the whole process.

An example of a questionnaire used in the evaluation is provided in Appendix C. The questions asked related to: how good they thought the computer application was to practise phonics; would they like to use it in school regularly; do they have access to a computer at home; usability of the sequencing activity; usability of the segmenting activity; would they like it on their computer at home. A summary of the mean responses to the questions is presented in Figure 6-3.

The maximum score is 5 (one for each smiley face from left to right) or a percentage where appropriate. The results are all positive (as is often the case with young children) but disappointingly only 60% would like the application at home yet all the children wanted to use it at school. On further questioning, it became apparent that the children associated the application as a learning tool for use at school and only wanted to play games or go on the Internet at home.

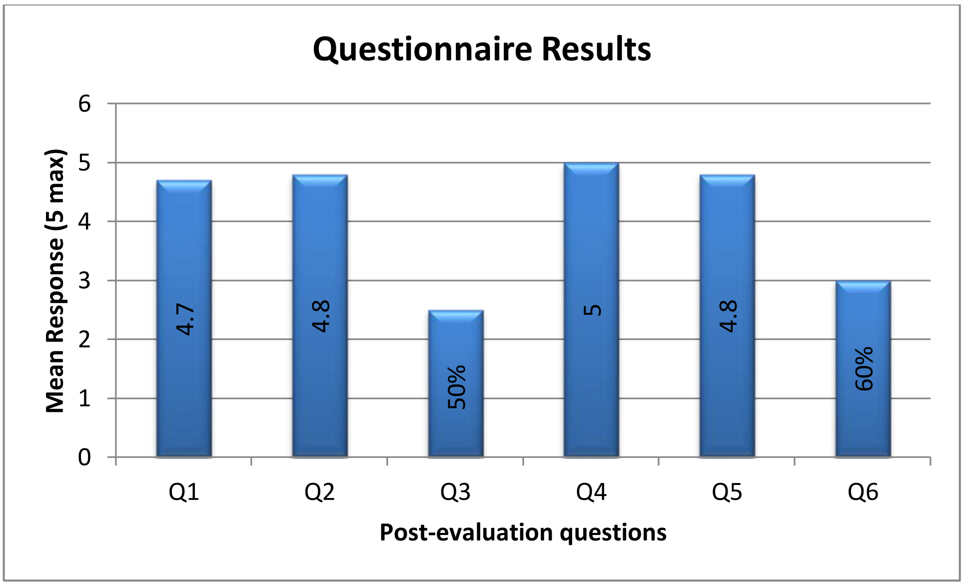

Figure 6-3: SPEL questionnaire results

6.2.4 Observational findings

Observational analysis during segmenting and sequencing sessions revealed some interesting points:

1. There were only two girls in the intervention group who clearly understood the concept of phonemes from the outset but one made significantly better progress than the other. When questioned, one of the girls (girl A) said she read every night with her mother but the other (girl B) said she did very little reading out of the classroom. However, girl A, who read a lot at home, was struggling with some of the more complex sequencing; it was postulated that she may have memorised the words by sight based on her increased reading time with her mother whilst girl B seemed to be logically working out the phoneme patterns. To test this notion, a nonsense word list was loaded into the SPEL application and both girls were asked to complete some sequencing and segmenting activities. The results showed that girl B got almost all the nonsense words correct, whilst girl A got hardly any correct despite her extra reading at home. Interestingly, the words that girl A was trying to sequence were real words that she knew from memory. This test is hardly conclusive and was carried out as a matter of interest whilst the children and the application were available. However, whether some children learn naturally

by memory because they lack natural logical ability, may be worthy of investigation and is suggested as an area for further work.

2. Share (2004) indicated that letter-name knowledge has a significant impact on letter-sound learning, even the letter names representing the corresponding long vowel sounds (A, E, I, O, U). The most common error thrown up by the observational results seemed to be the inability of some children to distinguish between the vowel digraphs that had an equivalent letter name, for example, when the children were expected to choose the phoneme vowel digraph <ai> as in "w**ai**t", most children chose <a> as in cat. All the children apart from girl B discussed in the previous section made this error at some point. All the children in the group who used the computer eventually rectified this error once they had made the sound↔symbol correspondence by clicking on the graphical letters on the screen and listening to the sound.

3. No time limit was imposed on the activities and no feedback was given, so the children did not mind clicking around the graphemes to learn the corresponding phonetic sounds.

6.2.5 Usability Conclusions

- The automatic help system providing instructions, choice-feedback and proactive help proved to be a hindrance and the usability was improved without it because the children were better able to explore and familiarise themselves with the sound↔symbol mapping. This does not mean that a help system should not be incorporated to avoid a deadlock situation where a child waits for the computer to do something and the computer waits for the child to do something. A well designed help system could overcome this type of situation by prompting the child after a given time period. The subject area of effective help for young children has proved to be a very interesting and challenging area of research and would make a research project in its own right. It is for this reason that this subject has been left as an area of further work which will benefit from the development carried out in this project and the lessons learnt from it.

- The children liked the screen help character “Floppy dog” as they had the choice of invoking this type of interactive help and understood his role as a help character.

- Most of the children found the graphics used for the graphemes colourful and large enough to click; three children thought the programme was not very exciting.

- All the children said they would like to use the application in the classroom and felt it would improve their phonics skills (the term phonics was used when discussing the application with the children, as they would not understand the term phonemic awareness).

- The Child-Computer Interaction guidelines were followed during the design of this application and were found to be effective in that the interface worked well. It is anticipated that the guidelines will be further refined through the development and evaluation of more computer applications.

- The action tracking facility was useful as it enabled sessions to be played back for analysis when details were not available from the video recordings because the position of the child in front of the camera sometimes obscured the view.

The interface usability was evaluated in a school setting, modified and re-evaluated until there were no new usability issues detected; it was now clear that usability issues would not impact on the RCT experiment documented in the following section.

6.3 Pragmatic Randomised Controlled Experiment

To enable a scientific approach to determine the effectiveness of the SPEL application an RCT approach has been used; this enables a quantitative analysis which is repeatable. This experiment aims to determine whether the SPEL computer system can develop young children's phonemic awareness skills in the classroom as effectively as a teacher. A neutral or positive result in favour of the intervention group would suggest that SPEL is a useful teaching tool which could reduce the staff resource intensive teaching of sound↔symbol mapping.

The null hypothesis to be tested is:

"The phonemic awareness improvement when using the computer will be the same as the improvement in performance from phonics teaching in the classroom."

6.3.1 **Pilot study summary**

A pilot study was carried out with year 1 pupils in a North West school. For the pilot study, the whole class of nineteen children were involved. Using a matched pair design (matched on age and gender), nine children were randomly allocated to a control group and ten children were randomly allocated to the intervention group. A similar approach was used in the main RCT and is discussed in detail in Section 6.3.2.3. The experiment aimed to establish whether children using the SPEL computer program would fare as well as those who had received the equivalent time by traditional classroom teaching. Children were given an author defined phoneme-to-grapheme and grapheme-to-phoneme correspondence test before and after a 10 week period during which they carried out activities on SPEL requiring them to build up and break down words. The pre- and post-tests used the same 36 question test where one mark was awarded for each correct answer. Differences in pre- and post-test scores were then analysed; the test results illustrated in Table 6-1 and Table 6-2 show that there was a small positive effect in favour of the intervention group.

	Group	Mean	Std. Deviation	N
PGbefore	c	24.0	5.0	9
	e	22.0	8.8	10
	Total	23.0	7.2	19
PGafter	c	29.8	3.6	9
	e	31.1	4.7	10
	Total	30.5	4.1	19

Table 6-1: Phoneme-to-Grapheme descriptive statistics

	Group	Mean	Std. Deviation	N
GPbefore	c	25.3	5.6	9
	e	27.4	5.5	10
	Total	26.4	5.5	19
GPafter	c	27.7	3.2	9
	e	31.7	4.0	10
	Total	29.8	4.1	19

Table 6-2: Grapheme-to-Phoneme descriptive statistics

In the phoneme to grapheme test the control group had a pre-test mean of 24 and the experimental (intervention) group had a mean of 22 with a slightly wider standard deviation indicating that the groups were of a similar standard before the experiment but the range of ability in the experimental group was slightly wider.

The before and after scores for the phoneme to grapheme test show an improvement of 6 marks for the control group and an improvement of 9 marks for the experimental group giving a performance difference of 3 out of 36 marks in favour of the experimental group. However, a statistical analysis using a one way ANOVA showed that the difference could not be classed as significant at the 5% level ($p > 0.3$) so these results could have occurred by chance.

In the grapheme to phoneme test, the control group had a pre-test mean of 25.3 which is comparable to that of the experimental group's mean of 27.4, with almost identical standard deviations confirming the equivalency in ability of the two groups.

The post-test gain of the control group of 2 marks is less than the experimental group's post-test gain of 4 marks. Although there is a performance difference of 2 out of 36 marks in favour of the experimental group, a one way ANOVA again showed the difference to be not statistically significant at the 5% level ($p > 0.29$) so this too could have occurred by chance.

Given the small number of children involved in the pilot, the gain was not shown to be statistically significant at the $p < 0.05$ level but it did however enable the SPEL application and research process to be tested before conducting a large scale experiment which would provide more accurate results.

The small standard deviations and relatively high means measured in the pre- and post-tests indicate that children were not stretched; the raw results show that the bulk of the children were able to handle most of the single character phoneme / graphemes. Any future tests would need to consider this ceiling effect.

6.3.2 **A large scale RCT**

A new experiment on the final version of SPEL was carried out using the same experimental process as the pilot with the exceptions of:

1 A power calculation was carried out to determine the number of subjects needed for the experiment to provide enough statistical power to carry out meaningful statistical analysis.

2 The pre- and post-tests were changed. A recommendation from the pilot suggested a different approach to testing to ensure the children are stretched and Brooks (2007a) recommends the use of a standardised test. The original 36 grapheme↔phoneme mapping questions were replaced by a standardised test to measure phonemic awareness skills. The "York Assessment of Reading for Comprehension (YARC) Early Reading" test is discussed in Section 6.3.2.5.

6.3.2.1 Power calculation for the SPEL randomised controlled trial

It is necessary to estimate the number of participants required to provide enough statistical power to enable meaningful statistical analysis of the RCT and to ensure resources are not wasted by including more participants than are required.

Cohen (1992) discusses at length the importance of, yet lack of, power analysis in many experiments and the apparent disregard of it by editors and reviewers. In an attempt to promote power analysis, Cohen goes on to provide "rule of thumb" values for effect size that enable an initial power analysis to be undertaken to estimate the number of subjects required in an experiment. Cohen states the four variables of statistical inference: "*sample size (N), significance criterion (α), population effect size (ES) and statistical power*." As each is a function of the other three, to determine N, the statistical power, significance criterion and population effect size need to be known or estimated.

In the case of the SPEL RCT, it was necessary to estimate the required sample size before the experiment started. The significance criterion is typically set to α=0.05 and the power is typically set to 0.8 (Cohen, 1992) so an estimate of the effect size was required. Cohen suggested three values for effect size: small, medium and large and allocates values to each based on the type of test used.

> *"The ES index for the t test of the difference between independent means is d, the difference expressed in units of (i.e., divided by) the within-population standard deviation. For this test, the H_0 is that d= 0 and the small, medium, and large ESs (or H,s) are d = .20, .50, and .80."*
>
> Cohen (1992)

Using the table in Cohen's paper (1992) it can be seen that for the mean difference for a two group ANOVA (or t-test), the number of subjects in each arm is 64 when α=0.05, power=0.8 and the effect size is medium (0.5). Using a power calculation tool (Faul et al., 2009), this value is confirmed at a total sample size of 128 illustrated in Figure 6-4.

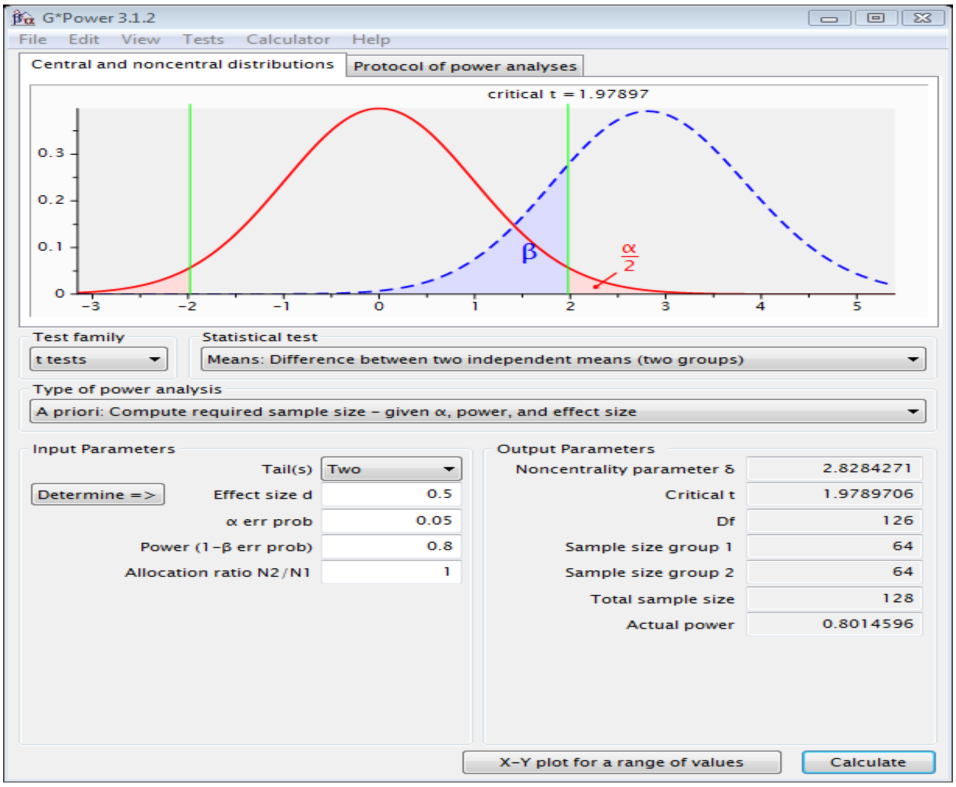

Figure 6-4: Power calculation results from G*Power

The G*Power computer program is a "*general stand-alone power analysis program for statistical tests commonly used in social and behavioural research*" (Faul et al., 2007).

The sample size of 128 (64 in each arm) assumes the sample is from a single population. It is highly unlikely that a single class within a single school will contain 128 year 1 children. A more typical class size in UK primary schools is 25-30 children. This would imply that about 5 classes of children would be required. This however presents a problem; although all the children will be following the same teaching scheme, each class is likely to have a different teacher and may be from different schools in different areas depending on availability. This clustering can reduce the statistical power of the test due to intra-cluster correlation effects. In order to maintain the 80% power, one of the other variables (α, N or ES) needs to be increased. As α is typically fixed at 0.05 and ES is typically fixed at 0.5 in trials of this type, this leaves N as the only variable. Killip et al (2004) explain that the effective sample size (ESS) is the result of dividing the single population sample size by a constant and therefore reducing it by a factor DE which denotes the design effect:

$$ESS = \frac{mk}{DE}$$

$$DE = 1 + \rho(m - 1)$$

Where: **DE** is the Design Effect

m is the number of subjects in a cluster

k is the number of clusters

ρ is the intracluster correlation coefficient

In this case, m and k have been estimated at 30 and 5 but the value of ρ needs to be estimated also. Killip et al (2004) claim that ρ is typically between 0.01 and 0.02 in human studies. Underwood et al (1998) state that many trials do not record the value of ρ but go on to say that in studies reported in General Practice, ρ is usually between 0.01 and 0.05.

As the SPEL experiment is a human study and erring on the side of caution, the higher figure of Killip et al and the centre of the extremes cited by Underwood et al will be used as guidelines for the estimate; ρ is therefore estimated at 0.03.

Rearranging the equation for ESS to set the number of clusters as the variable:

$$\frac{ESS\big(1+\rho(m-1)\big)}{m}=k$$

Where: **ESS** is calculated by G*Power to be 128

ρ is set to 0.03

m is 30 for a typical UK primary school class

$$\frac{128\big(1+0.03(30-1)\big)}{30}=7.98$$

So with a class size of 30 the experiment will require 8 cluster or school classes of 30 pupils so a sample size of 240 in total is required. However, the classes cannot be guaranteed to be exactly the same size (but they should be close) which further slightly weakens the power (Eldridge et al., 2006). There may be attrition due to some children leaving the school during the experimental period so to provide headroom to protect the statistical power nine classes of thirty will be used in the SPEL RCT; the total number of participants is therefore estimated to be 270 children – 135 in each arm.

Using the new cluster value, the new ESS can be calculated:

$$ESS = \frac{mk}{1 + \rho(m - 1)}$$

$$ESS = \frac{270}{1 + 0.03(30 - 1)}$$

$$ESS = 144$$

Using this new potential ESS in G*Power provides a power value of 91%. So using the sample size of 270 over 9 clusters will ensure a statistical power between 80% and 91% depending on how closely matched the cluster sizes are in practice.

6.3.2.2 Experiment design detail

A pragmatic RCT was conducted in a classroom setting of year 1 pupils aged 5 to 6 years. 266 children from four schools in the North West of England were involved in the trial: two three-form entry schools (A and H); one two-form entry school (R) and one single-form entry school (L). Typically, OFSTED describe rates of free school meals (FSM) using the terms "none", "all", "above average", "average" or "below average" (NUT report, 2009). In terms of the FSM OFSTED ratings, the provision of free school meals is commonly used as an indicator of socio-economic status. The school participants as a sample based on this indicator are: school A – below average; school H – above average; school L – below average and school R – above average. The class sizes in Table 6-5 show schools to have similar class sizes to those based on the power calculation carried out in Section 6.3.2.1; this indicates that the power will be in excess of the minimum target stated at 80%.

Randomisation of control and intervention subjects avoids selection bias (as discussed in Section 4.2.4) on the part of the teacher, who may place children in a particular group because they have prior knowledge of the children's behaviour, for example. For the RCT to be effective, as many variables as possible needed to be controlled (other than the independent variable of intervention). Based on the discussion of age and gender as potential confounds in Chapter 2, the experiment did not discount gender bias (discussed in Section 4.2.2) and age bias (discussed in Section 4.2.3). The randomisation process if carried out effectively can control the variables of: school and class within school (one school may be more effective than another, one teacher may be more effective than another); age and gender of the children. For example, randomising on matched pair (by age and gender) within class automatically controls for school and class within school as well as age and gender. Random allocation was therefore carried out on pairs matched on age and gender at a class level to ensure that any differences between classes, schools, age and gender were diluted in the overall results as this method ensured that almost exactly half of each class, school, age and gender would be in the intervention group.

A 'waiting list' design was adopted for the experiment to: avoid the demoralisation effect (discussed in Section 4.2.8) and to attempt to reduce the rate of attrition in order to avoid attrition bias (discussed in Section 4.2.10). To avoid this type of bias, the SPEL application was left in the school for them to keep and the control group were made aware prior to the experiment starting that they would get the chance to use the computer system following the experiment to avoid them feeling excluded and demoralised and also to avoid them asking to be moved over to the intervention group to use a computer during experimental sessions. Either of these occurrences would have caused a problem for the teachers who had agreed to ensure that children stayed in the groups to which they were originally randomised; the waiting list design appeared to have worked, since teachers reported no such problems.

A great deal of effort had gone into controlling the potential "usability bias". Extensive research had not found another experiment in this domain that took this bias into account. If the computer application had been difficult to use, this may have introduced dilution bias (discussed in Section 4.2.7) to the intervention group and the results would have been likely to be skewed in favour of the control group. The phonemic awareness computer application which was used by the intervention group had been custom built and evaluated using qualitative research methods (discussed in

Section 4.3) in the classrooms of schools not involved in the experiment to ensure that the interface was simple and intuitive.

ITT analysis (discussed in Section 4.2.4) dictates that analysis of data is carried out on the original randomised groups; once allocated to the control or intervention group, participants must not change groups once the experiment is underway. ITT analysis avoids selection bias which may occur if children were to ask to swap groups and teachers were to allow them to do this. Once this experiment was underway, all participants remained in their allocated group (the teachers were made aware of this requirement and signed an agreement to that effect).

Chapter 2 discusses the lack of UK based rigorous studies undertaken to evaluate the effectiveness of ICT for phonemic awareness, finding only one RCT which evaluated the effect of ICT on reading and spelling scores in older children aged 11-12 years; no UK study of this kind for children aged 5-6 with a sample size large enough to enable reliable results was found in the literature. The rigorous scientific approach detailed in this document will ensure a significant contribution to the research community. The results of this RCT should also be useful to researchers carrying out meta-analysis studies.

6.3.2.3 Method of randomisation

A spreadsheet was created to hold the following information in columns:

Child ID: a unique ID provided by the school. Names were not recorded.

Date of birth

Gender

Months: a calculation to calculate a child's age in months without rounding (non-rounding is specified in the YARC data section of the manual).

Difference: This column calculated the difference in age of adjacent rows. This simplified the identification of the closest age matches.

Group: this contained either "c" or "i" to indicate allocation to control or intervention.

Pair: this used a character to clearly indicate which two children have been paired.

Randoms: the actual random value used to allocate the control or intervention choice.

Random Booleans: a list of random Boolean values generated by the algorithm detailed in Appendix F.

The school office populated the spreadsheet with Child ID, Date of Birth and Gender. In order to avoid selection bias (as discussed in Section 4.2.4) on the part of the teachers and the researcher, who may allocate children to particular groups through personal preference, a third party was given the spreadsheet and carried out the randomisation using the following process:

1. Sort each sheet by gender: males and females will be blocked with females in the first half of the sheet and males in the second.
2. Sort the girls by age.
3. Sort the boys by age.
4. Using the age difference column, identify suitable female pairs and label them. Generally these will be adjacent pairs as the children have only a possible 12 month age difference. If a child is more than 2 months different in age than the adjacent child yet the following adjacent pair is not, then miss out the child with the large age difference and move on. This will ensure that the majority of pairs will have the smallest age difference.
5. Do the same for the males.
6. There are a number of possible scenarios after the first pass. These need to be dealt with sensibly. For example, there may be a single outlier left in the boys and another in the girls. If so, pair these. If there are more outliers then pair on the nearest age matched by gender where possible. If there is a single child left unallocated, they will be randomly placed in control or intervention. This method will ensure that the effect of outliers is reduced when taken across the whole sample.

   Any pairs allocated that are not adjacent in age and gender should be marked-up in the spread sheet with an explanation of how they have been allocated for audit and repeatability.
7. Identify the first pair. Using the first random number in the sequence, allocate the first child in the pair to 1=Control or 0=Intervention then allocate the second child to the opposite group.

8 Calculate the mean and standard deviation of the control and intervention group ages to provide an indication of how evenly the groups have been allocated on age. However, regardless of the distribution, the randomisation will not be repeated; if the randomisation process has worked effectively, these figures should simply confirm a reasonable distribution. If the distributions are very different, the randomisation process is likely to be flawed so the whole randomisation approach would need to be revised.
9 Repeat the process for each class in each school.
10 Send back the list of control and intervention group child identification number to the class teachers. The teachers will use this list to allocate children to the control group or SPEL intervention group for the twelve week experimental period.

6.3.2.4 Experimental process

Torgerson and Torgerson (2007) note the importance in educational experiments that the intervention is shown to work in a usual school setting rather than an artificial setting with poor external validity. The RCT was run over a 12 week period, during which time the intervention group used the computer application in the school computer suite for half an hour a week whilst the control group were taught the same symbol↔sound mapping using the same words by traditional (paper-based) teaching methods in the classroom. All pupils in the study still continued normal classroom teaching of the primary curriculum which includes a statutory daily phonics session; the control and intervention exercises were in addition to this.

Contamination bias (discussed in Section 4.2.5) can arise in a number of ways where interaction between experimental and control groups can invalidate direct comparisons. For example, the control group could learn information intended for the experimental group if the control group could see or hear the computer application or the control group could gain access to the program. To avoid contamination effects between the intervention and control group, the intervention and control group sessions took place in different rooms of the school and teaching staff were asked to ensure that the control group children did not use or see in use the application during the experimental period and signed an agreement to this effect. The computer technicians for each school were also asked when they installed the application not to provide a short-cut to the application through a desk-top icon; the application was accessible only through the Windows programs menu at: start | all programs | UCLAN | SPEL so it is not a simple path for a child to guess; staff running the computer intervention group sessions in the computer suite always set the program running directly prior to each intervention computer class in order that the children in the intervention group did not become accustomed to starting the application and pass this information on to children in the control group.

6.3.2.5 Tests

The age-appropriate UK developed test chosen to measure performance in phonemic awareness was "The York Assessment of Reading for Comprehension (YARC) Early Reading"; a suite of standardised, paper-based reading assessments for use in primary schools developed with experts in reading at the University of York (Snowling et al., 2009). The YARC test was used to collect phonemic awareness performance data on all children involved in the experiment both before and after the 12 week experimental period. The test provides the option to convert raw data into standardised scores, which can then be compared to norms, constructed by the test authors from a fully representative standardised sample, covering the age range 4 years to 7 years and 11 months.

The YARC test package was purchased to carry out pre- and post-testing. However, the test results need to be calculated manually by table look-up in the reference booklet. Given the large number of participants, this needed to be automated. A Microsoft Excel workbook was created with all the table details and functions were created to carry out the look-up operations and present the results within the workbook. From the many worksheets within the workbook, reports for the teachers were created. The relevant data was extracted into separate sheets in a format suitable for entry into the statistical analysis package SPSS.

An example of a test sheet derived from the YARC test pack is illustrated in Appendix D; this was designed to combine all tests onto a single sheet for efficiency and consistency of data collection. The YARC test manual recommends a combined score from the Sound Isolation and Sound Deletion tasks to be presented as an overall phonemic awareness score in educational studies. Given the effort involved in setting up the RCT with such a large sample it was considered to be a valuable opportunity to collect other data at the same time:

- data was collected for the full suite of YARC tests.
- an author defined split vowel test (illustrated in Appendix D) was carried out to avoid the problem caused by test ceiling effects encountered in the pilot study discussed in Section 6.3.1 should this have happened in the main experiment. Fortunately, ceiling effects were not found to be an issue in the main RCT and therefore these test results were not required.
- a digit span test (Koppitz, 1970) (illustrated in Appendix D) was also carried out as a result of a conversation with Professor Rhona Johnston from the University of Hull who stated that a comparison of performance between phonemic awareness and digit span may turn up some interesting correlation conclusions for researchers interested in this area.

Collected data can be provided on request by contacting the author via the School of Computing, Engineering and Physical Sciences office at the University of Central Lancashire.

The results sheets contain data from several tests but only the summed Sound Isolation and Sound Deletion raw scores are used in this RCT as they are the YARC test indicators of phonemic awareness.

Data collectors were recruited for the experiment and were trained in the administration of the tests and recording of data. All children were given the same pre- and post-test directly before and directly following the 12 weeks experimental period. The tests were carried out in a quiet area of each school.

To ensure blind assessment of outcome regarding the pre- and post-tests, the data collectors did not know whether the child they were testing was in the intervention or control group and the teachers did not give any indication of the group membership when each child was sent to be tested; the teachers signed an agreement to this effect.

This controls for reporting bias (as discussed in Section 4.2.9). The data collectors were sent a spreadsheet containing a list of user ID numbers sorted numerically to ensure that the control and intervention groupings were invisible to them. The data collectors input the test scores onto this spreadsheet. Summing, collation and data input to an Excel spreadsheet was cross-checked for accuracy by the data collectors. Copies of the paper test sheets have been kept for experimental audit purposes. One final step before the analysis was carried out was the cross-checking by a data collector of every entry in the data analysis table against the hand recorded paper copies to ensure the data had been accurately managed. The results are reported in Section 6-30. A full audit trail of the process was maintained to ensure that the researcher had no opportunity to influence the results.

Hanna et al (1997) found that young children could concentrate for approximately 30 minutes. The four tests from the YARC Early Reading series take approximately twenty minutes to complete (Snowling et al., 2009). The digit span test and split vowel test took an additional 5 to 10 minutes to complete.

The phonemic awareness tests relevant to this project comprise the YARC Sound Isolation and Sound Deletion tests and took approximately 5 minutes each. The tests and questions can be asked in any order but for consistency the tests were administered in the order found in the YARC test booklet: Letter Sound Knowledge, Early Word Recognition, Sound Isolation and Sound Deletion. The user defined tests of Split vowel and Digit span completed the test session which lasted approximately 30 minutes in total. Questions were asked in the order as listed on the test record sheet (Appendix D).

The YARC test instruction manual (Snowling et al., 2009) details the approach and was strictly followed by all data collectors to ensure consistency. It would be necessary for other researchers to purchase the manual if the experiment was to be replicated. However, a summary of key elements to the test approach for the phonemic awareness tests is:

- Test questions with full feedback are provided for practice.

- Repeat question once if required.
- One mark is awarded for a correct answer; zero is awarded for an incorrect answer or no response.
- Prompt once only for an answer if there is no initial response.
- No feedback or assistance is given either directly or indirectly (through gesticulation, facial expression or sound).

6.3.2.6 Intervention and control treatment

The children in the intervention group used the SPEL application for 30 minutes, once a week for the twelve weeks between the pre- and post-test. The intervention group were supervised by staff trained to use the application by the developer of SPEL; although it is very simple to use, there are advanced override facilities should there be any unforeseen issues with the application (discussed in Appendix E). However, no issues were found with the application from any of the schools. In the weeks between the pre-test and post-test, the children in the intervention group used the SPEL application to sequence and segment words with new graphemes introduced progressively based on the "Letters and Sounds" reading scheme recommendations (DfES, 2007b) ranging from simple to more complex CVC words through the introduction of consonant and vowel digraphs and finishing with split vowel digraphs. The activities are detailed in Section 5.4.2.

Traditional paper based phonics sessions were delivered to the control group at the same time and for the same duration each week whilst the intervention group received their computerised phonics sessions. A paper version of the content of the weekly computerised phonics session, with regard to graphemes, phonemes and words to be taught was given to the teachers prior to each session so that the only difference during the learning sessions between the groups was the mode of delivery.

Treatment of the intervention and control groups in this way avoided dilution bias (discussed in Section 4.2.7), which can occur if the control treatment is not delivered adequately, which may bias the experimental results by affecting the performance of the control group in a negative way. Asking the teachers to mirror the content of each intervention session reduced the possibility of the control group being disadvantaged.

6.3.2.7 Word Lists

Word lists and activities (segmenting or sequencing) can be changed at predefined dates or manually. Appendix E provides full details on operating the application and explains how to exploit its flexibility if required. For the purpose of the experiment, the application was configured to alternate from segmenting to sequencing with different word and grapheme sets on a weekly basis. The grapheme and phoneme sets and the word lists used are detailed in Appendix B; the phonemes listed in the tables in Appendix B are represented by the International Phonetic Alphabet symbols (IPA., 2005).

The lists start with simple, familiar words containing simple graphemes building up to less familiar words which contain the more complex graphemes. The word sets ensure that each grapheme is covered several times over the course of the 12 week experimental period.

The word lists have been drawn primarily from the DfES Letters and Sounds scheme (DfES, 2007b), but to increase the number of words available to the activities, further words have been provided by a school Literacy Co-ordinator. The word lists have been designed to avoid floor and ceiling effects; there are some simple words and there are enough words in each activity to ensure the children do not finish an activity within the half hour. Each set of digraphs is covered completely before any elements are repeated; for example, if the activity uses six digraphs, then words using each digraph will be presented to the children, then the set of digraphs is used again but in a different order and within a different set of words.

6.3.3 SPEL Pragmatic Randomised Controlled Trial results

The results of the pre and post tests were analysed using SPSS v16. The analysis approach is to compare the difference in means of the pre and post test results then determine whether any difference is statistically significant at $p < 0.05$. The key outcomes are reported here but the detail behind the results is presented in Appendix H.

The means of the groups are illustrated in Table 6-3.

Report

GroupedPost		Pre Test Results	Post Test Results
Control	Mean	13.6	16.5
	N	133	133
	Std. Deviation	6.0	6.0
Intervention	Mean	13.5	16.3
	N	133	133
	Std. Deviation	5.6	5.3
Total	Mean	13.6	16.4
	N	266	266
	Std. Deviation	5.8	5.7

Table 6-3: Mean and standard deviation of each group

The pre-test control and intervention groups are less than half a mark different on pre-test mean results with similar standard deviations which indicates that the two groups were well matched. The potential covariates of age and gender were accounted for in the randomisation process before the pre-test so to minimise the potential covariate effect of ability. An analysis of covariance (ANCOVA) was carried out using pre-test scores as the covariate. The results of this test (detailed in Appendix H) show that the difference in post-test performance between the control and intervention groups is not significant ($p > 0.7$).

It is interesting to note from Table 6-3 that the change in pre- and post-test means for both control and intervention groups is about 3 marks which indicates that the phonemic awareness has improved for both groups over the twelve weeks of the experiment. A paired t-test was used to test the statistical significance of this increase. As both pre-test and post-test groups have been shown to be similar, the t-test was used to compare the mean of all post-test results to the mean of all pre-test results.

The result of the t-test of the whole group shows the 3 mark increase to be significant $p < 0.001$. This result suggests that the computer group progressed at the same rate as the taught group and each group progressed with a positive medium educational effect size (d = 0.5). Details of the calculations of Cohen's *d* for these results can be found in Appendix H. The results imply that the SPEL system is likely to be as effective as extra teacher led classes and is therefore likely to be a useful teaching tool for phonemic awareness practice.

6.3.3.1 Attrition

Out of the initial 266 children, 7 children left school before the experiment was completed. To comply with the ITT approach (discussed in Section 4.2.10 - *Attrition bias*), all results were submitted for analysis; results for the missing children were recorded as zero. The attrition profile is summarised in Table 6-4.

School	Intervention		Control		Total
	Male	Female	Male	Female	
A	0	0	0	0	0
H	1	1	0	0	2
L	0	0	0	0	0
R	1	0	3	1	5

Table 6-4: Attrition profile

The total attrition: 7 / 266 or 2.5%

Intervention attrition: 3/133 or 2%

Control attrition: 4/133 or 3%

Differential attrition: 1%

Attrition as low as this is unlikely to affected the results (discussed in Section 4.2.10 - *Attrition bias*). However, for completeness, the missing children's results were set to the extreme value of full marks (24/24) and the tests were re-run; results are provided in Section H7 - *Attrition test*. There was a small change in means but the difference was still not statistically significant ($p > 0.69$). The results are detailed in Appendix H - *Statistical Analysis Details*.

6.3.3.2 Summary of results

The following table summarises the results of the RCT. The summary items are based on recommendations from Brooks (2007b) to enable enough information for others to include the results, for example, in a meta-analysis, experiment comparison or replication of results. Each school is depicted by the first character of its name: A, H, L or R. Each result is cross-referenced to the detail where appropriate.

Requirement	Response
Name of intervention	SPEL (Chapter 5)
Main references	Included in this document (References section)
Research design	Pragmatic RCT. N=266 (Section 6.3)
Date when it was implemented	Pre-test:25/01/2010 – 05/02/2010 Post-test:31/05/2010 – 11/06/2010 Experiment period: 15/02/2010 - 21/05/2010 *Note: The experiment was run for 12 weeks; the fourteen weeks between the dates above include a two week break for Easter 29/03/2010 – 09/04/2010*
Age range of children	5 to 6 years (Section 6.3.2.4)
Type of children involved	Year 1 mainstream mixed primary (Section 6.3.2.4)
Number of schools (A, H, L, R)	4
Number of classes	A = 3: 30, 30, 30 = 90 pupils H = 3: 26, 27, 26 = 79 pupils L = 1: 37 = 37 pupils R = 2: 30, 30 = 60 pupils Total 266 pupils
Number of pupils in experimental group	133
Number of pupils in control group	133
Attrition *Based on intention to teach, all children have been left in for analysis even if the post-test was not administered as 7 children left school before the post-test*	A = 0 / 90 H = 2 / 79 (2 intervention: 1 male 1 female) L = 0 / 37 R = 5 / 60 (3 males control, 1 male intervention, 1 female control) Total 7 / 266 or 2.5% (Section 6.3.3.1)
Whether groups were equivalent	Intervention and control groups had the same number of pupils allocated randomly paired on age and gender. Mean and SD of the pre-test illustrates equivalence and graphs of normal distribution show the groups to be reasonably normal (Appendix H)
Length of intervention in weeks	12 weeks (Section 6.3.2.4)
Assessment	The York Assessment of Reading for Comprehension (YARC) Early Reading test. Sound Isolation and Sound Deletion raw scores were added to provide a single raw phonemic awareness score for each pupil in the range 0 to 24 (Section 6.3.2.5)

For each group, pre- and post-test means (rounded here to full marks but detailed in appropriate section)	Pre-test control: 14/24 Pre-test Intervention: 13/24 Post-test Control: 17/24 Post-test Intervention: 16/24 (Section 6.3.3)
Pre- and Post-test score differences	Both groups gained 3/24 (Section 6.3.3).
Effect size (Cohen's d using pooled SD)	Cohen's *d* = 0.5 pre- to post- test gain Cohen's *d* = 0 at post-test between the groups (Appendix H)
Statistical significance at the 5% level	Post-test groups difference: p > 0.7 Pre- Post-test gain: p < 0.001 (Section 6.3.3)
Progress summary	Both intervention and control group showed the same medium sized educational improvement over the experimental period (Section 6.3.3)

Table 6-5: Summary of RCT results

6.4 Conclusions

This chapter provides a detailed discussion of the research process undertaken to carry out a study into the effectiveness of the SPEL programme in developing the phonemic awareness of young children.

The first part of the chapter details the qualitative evaluation required to remove the application's interface usability as confounding factor in the RCT; this is a significant contribution to knowledge as no other studies have been found that report this potential confound.

The second part of the chapter discusses the randomised controlled trial pilot RCT; the pilot enabled a trial run and lessons learnt to be applied to a large scale RCT involving four schools and 266 children. All the children took a pre-intervention test to provide reference data. The intervention group used the SPEL program in a computer suite over a three month period for 30 minutes each week during which time the control group were taught the same sets of words and sounds using traditional teaching methods in the classroom. The 30 minutes each week attributed to the experiment was extra tuition the children would not normally have received so did not detract from their normal teaching schedule for either group. On completion the children were tested again (using the same test) and the difference in performance of the intervention group and the control group was measured. In an experiment of this nature it is necessary to keep all non-measurable variables under control – constant if possible.

To maximise the validity of the outcome, the confounding factors that were controlled and therefore minimised were:

- Randomisation of participants controlled for selection bias as this process was carried out by a third party who had no knowledge of the children.
- Age / experience and gender were accounted for by using a matched pair design for the random allocation.
- Ability was not initially controlled (other than by pairing on age and gender) as that would have required another test. However, the pre-test scores were used as a covariate in the analysis.
- A waiting list design controlled for performance bias of parents or teachers who may have otherwise provided extra tuition for children in the control group. This design also controlled for the demoralisation effect and attrition bias which may have been brought about by teachers changing the original randomised groups due to control group children who preferred to use the computer during the experimental sessions.
- Attrition bias was avoided by data collectors returning to school to carry out a 'mop-up' exercise for children who were absent during the testing period. Seven children left school during the experimental period; approximately 2.5% spread across both arms. The children's results were left in the data for analysis in accordance to ITT. As a rule of thumb, less than 5% attrition is not a problem (Torgerson, 2007).
- Reporting bias was controlled by recruiting data collectors who were blind to treatment allocation.
- An agreement signed by school staff: controlled contamination bias of the control group, which may have otherwise gained information about the

computer application; controlled selection bias of teachers who may have otherwise changed the original randomised groups and controlled reporting bias of the data collectors as this ensured that they were blind to the group allocation.

- A full audit trail of the data collection, randomisation and analysis is available for inspection to verify the integrity of the process.
- Application usability could be a significant factor – if the application is difficult to use, it could create a number of adverse effects such as confusion and frustration causing a negative learning experience. It is not clear why others have not included usability as a significant factor in their studies – perhaps it is because of the amount of work required or maybe it was not thought about or perhaps it is not considered to be a problem (but no evidence was found to support this). In this case every effort was made to rule it out by designing and evaluating a simple but effective user interface. A key benefit in controlling this variable has been the development of a set of interface design guidelines which have been published for the benefit of the research and development community. The lengthy period of the project spent developing the guidelines has enabled a significant level of experience to be gained in working with young children in the classroom evaluating interfaces.
- Development of a computer application tested for usability by a group in the target age range of experimental participants controlled for dilution bias of the intervention group, which may have otherwise been brought about by the intervention not being delivered adequately by staff supervising the computer sessions, which may have affected the performance of the intervention group in a negative way.
- Treatment of the control group avoided dilution bias in that teachers were given a paper-based version of each phonemic awareness session to be delivered by the computer; this consistent approach to teaching content of control group sessions minimised the chance of these sessions not being delivered consistently, which could have otherwise affected the performance of the control group in a negative way.

Although the results show no significant improvement in the learning rate when using a computer program even though it was designed carefully for this age group, the results do suggest that the intervention group developed at the same rate as the control group which received extra traditional teaching so the application could be used as a teaching tool to relieve the teacher or provide extra practice in phonemic awareness training in the classroom or at home.

The raw data collected under strict experimental conditions for a battery of tests is available as a contribution to the research community to carry out further work or replicate the results detailed in this report.

Although this study is one of the largest of its kind, the generality of the results is limited as the study was confined to four schools in the same geographical area.

The journey through this PhD project has generated many interesting questions, many of which are recommended areas of further work; these questions and proposed further work are included with the overall thesis conclusions in the next chapter: *Conclusions and Recommendations for Further Work*.

Chapter 7

Conclusions and Recommendations for Further Work

Chapter 7 Conclusions and Recommendations for Further Work

7.1 Thesis aims

The aims of this research project are summarised in this section for reference and the following sections discuss the extent to which they have been met.

- To establish the problems and benefits of using computers as teaching aids in the early primary classroom. Once known, the problems may be addressed and the benefits maximised.

- To understand the approach to teaching phonics in the early primary classroom to inform the design of a phonemic awareness tutor with an appropriate approach, scope, level and content.

- Produce and evaluate an intuitive user interface for young children to ensure that poor usability of the interface would not bias the results of an RCT in favour of the control group.

- Implement an effective educational tool for early primary classroom teaching designed around academic theory in the area of phonics teaching in the UK using a computer interface specifically designed and evaluated for the target age group.

- Carry out a randomised controlled trial to determine the educational effectiveness of the phonemic awareness tutor.

7.2 Conclusions

Each chapter contains conclusions of specific areas of work. However, this section summarises those conclusions in the context of the project's aims.

7.2.1 Problems and benefits of using computers as a teaching aid in the early primary classroom

This aim was addressed by means of a literature review documented in Chapter 2 - *Computers in the Classroom*. The conclusions are summarised below:

- Multimedia computers offer new and dynamic ways to learning through a range of media including graphics and sound compared to traditional text-based literacy learning.
- More quantitative evidence is required regarding the usefulness and effectiveness of ICT for reading and spelling in the classroom.
- There is a shortage of randomised control trials to evaluate the effectiveness of computerised phonics programs in UK primary classrooms, particularly for children younger than 7 years of age in mainstream education.
- There is evidence that many teachers are under-trained in the use of ICT.
- Teachers and developers tend to shoehorn existing teaching practice into the computer domain.

7.2.2 To understand the approach to teaching phonics in the early primary classroom

This aim was approached by means of a literature review documented in Chapter 3 - *Phonics in UK classrooms: the debate*. The conclusions are summarised below:

- The teaching of phonics has been particularly troublesome.
- Synthetic phonics surfaces as the technique of choice today.
- Synthetic phonics teaching should use a multi-sensory approach where possible.
- Consistent pronunciation of the phonics sounds is required for the technique to work properly.
- The teaching of phonics should be systematic, quick and early, as opposed to incidental, slow or late.

- Primary schools should use a recommended phonics teaching scheme such as Letters and Sounds (DfES, 2007b).

- The order in which the sounds in words are learnt should be first-middle-last.

- Gains could be made if technology was used effectively and supported by teachers.

7.2.3 Produce and evaluate an intuitive user interface for young children

This aim was approached by action research. A summary of the conclusions of this phase of work are:

- Through the development and evaluation of several literacy-based computer programs for young children, a set of child-computer interaction design guidelines was created which will be useful to the research field of Child-Computer Interaction (ChiCI). Details of this work can be found in Appendix A - *Child-Computer Interaction*.

- The System for Phonics Early Learning (SPEL) was designed and implemented using the developed guidelines and the evaluation results were excellent. That is not to say that they would work as well for all application types but it is a starting point for researchers and developers; the challenge and modification of the guidelines by other researchers is welcomed.

- One guideline suggests using pro-active interactive help. This was attempted in SPEL but was removed as it became a hindrance to exploration; it is clear that the design of an effective help system is a complex area and a research project in its own right.

7.2.4 Implement an effective educational tool for primary classroom teaching

The SPEL application was developed using the C++ programming language. Hundreds of speech files were recorded to produce random but appropriate concatenated human speech which provides a more natural interaction with the user. The application was packaged into a Microsoft .msi installation file set and made available from the Internet for schools to download. An overview of the SPEL application is provided in Chapter 5 - *SPEL Operation and Interface Design* and its extensibility is detailed in Appendix E - *SPEL Application Configuration*. A summary of the conclusions from this phase of work is:

- A tool was produced and used over a substantial time frame by 19 children during a pilot RCT and 266 children during a large scale RCT; no programming errors occurred and no usability issues were reported during this period.

- The application provided sequencing and segmenting exercises. The computer was able to make the phonetic and complete word sounds which enable children to learn the sound symbol↔mapping. The sequencing and segmenting exercises enabled the children to appreciate that words are made up of groups of symbols and associated sounds. The application is flexible and extensible to enable phonically decodable words and sounds to be used.

- SPEL has been shown to be usable by children as young as 5 years with minimal help. This frees teachers' and teaching assistants' time.

- The teachers who observed the evaluation sessions were delighted with the potential of SPEL as a phonemic awareness tutor as most children quickly grasped that the symbols on the screen were associated to the sounds in words;

even incorrect choices acted as a teaching mechanism, as the children could learn, through exploration and general feedback, the sound↔symbol correspondences with each click of a grapheme.

7.2.5 Carry out an educational effectiveness RCT

Details of the design components of this phase of work are available in Chapter 4 - *Research and Evaluation Methods.* Implementation details and summarised results of this work are reported in Chapter 6 - *SPEL Usability and Phonemic Awareness Experiments*. Detailed results can be found in Appendix H - *Statistical Analysis Details*.

- A pilot RCT was carried out in a single school by the whole of year 1 pupils (N=19). The results showed a very small improvement in favour of the intervention group but this gain was shown to be not statistically significant so could have occurred by chance. To increase the statistical power to determine whether there was an improvement in favour of the intervention group, a larger scale RCT was carried out.

- An RCT involving 266 year 1 children from 9 classes in four schools was carried out. The results showed no statistically significant advantage to either group in the learning rate of phonemic awareness. The intervention group using the computer program and the control group using the traditional teacher-delivered paper-based approach progressed at the same rate.

- The results did suggest however, as the intervention group developed at the same rate as the control group, that the SPEL computer program could be used as an effective teaching support tool which can reduce the amount of resource intensive tuition required by children in this age group.

7.3 Thesis conclusion

Can computers be used to develop phonemic awareness in the early primary classroom?

This research indicates that phonics lends itself well to computerisation. Interface design is particularly important for applications designed for young children. A system was built with a suitable interface and testing suggested that it appeared to be at least as effective as the phonics tuition provided by the classroom teacher. The application could therefore be used as a useful teaching aid.

There were no technical or usability issues reported from any of the schools so it therefore does appear that computers can be used to develop phonemic awareness in the early primary classroom.

7.4 Original contributions to knowledge for the research community

- An urgent need to undertake a pragmatic RCT to establish the effectiveness of a computerised phonemic awareness tutor in UK classrooms for children aged 5 to 6 years was identified in Chapter 2 - *Computers in the Classroom*; no such study was found, with sufficient statistical power to achieve statistical significance, prior to this experiment being carried out. The findings from this study will be useful to government bodies, educationalists and researchers involved in making recommendations for future educational policy regarding the introduction of computer software for phonemic awareness into mainstream schooling.

- A clear explanation of the RCT process and analysis of results should be of value to others wishing to undertake such a trial.

- Although only phonemic awareness data was used in the analysis, due to the magnitude of the task of setting up an RCT of this size, data from a battery of literacy-based tests was collected during pre- and post-testing; this data is available on request.

- The pragmatic nature of this experiment provides a contribution to knowledge because the intervention was delivered in the school computer suites with minimal staff supervision which provides a true reflection of the resource costs of adopting the intervention now it has been established as a useful teaching tool.

- The nature of the experiment design, following the guidelines for RCT's, facilitates replication and assures the validity and integrity of the results for use in this or similar areas of educational research.

- The rationale behind the choice of phonics teaching in UK classrooms is provided in Chapter 3 - *Phonics in UK classrooms: the debate*; this provides a list of the key areas of consideration for future designers of both computerised and non-computerised phonemic awareness programmes for primary classrooms in the UK.

- A set of interface design guidelines has been stabilised through application development and evaluation experience with child-evaluators in the classroom. There is surprisingly little research in this area; this work makes a contribution to the research area of child-computer interaction.

- Several peer reviewed research papers have been created and presented at international conferences; these are cited throughout this document.

7.5 Recommendations for further Work

- SPEL is a prototype designed specifically for this research project. There is evidence that the application is effective in developing phonemic awareness and the teachers found it a useful tool in its prototype state; this suggests that it would be useful to complete the application by developing an intuitive supervisor interface from where the teacher or parent could set various user options and preferences. This is where a voice recording facility would be made available. However, as teachers and parents are not necessarily computer experts or even interested in them beyond any teaching benefit they can bring, it is essential that any supervisor's interface should be designed with simplified usability and navigability or it simply will not be used. This however is not research work; a programming undergraduate for example could complete this application as their final year project with guided supervision.

- The design guidelines need to be extended into domains other than literacy to improve their generality. Input from other researchers is needed to improve and develop them into a more robust and generally accepted set.

- There is scope to further explore tiered interactive help approaches. This is a research project in its own right. It would be a very useful feature to add to SPEL but it would be more important to create general guidelines on how to integrate various help components into an application such that it operates intelligently by monitoring the user interaction then adapt to the user's progress.

- Work on dynamically adapting to the child's progress would be very useful. If SPEL was to become an intelligent tutoring system, it should be able to monitor the child's progress and determine the next question or activity based on the current success rate or patterns of errors.

- A project to study the effects of regional accents on the delivery effectiveness of phonics would be useful. If a problem is discovered, perhaps children in some regions may be weaker in the area of phonics because of incorrect pronunciation or enunciation by the teacher, then at least the problem will be known and a solution could be sought.

- Working with young children over a number of years, watching them explore, solve problems and ask questions – some patterns of behaviour have been seen which may be worthy of further exploration. One in particular is where the two girls discussed in Section 6.2.4 - *Observational findings*, seemed to deal with the nonsense word test quite differently. One girl clearly used the segmenting and sequencing approach to try to deal with a word she had never seen before whereas the other girl tried to correlate the nonsense word with ones she knew. Is it possible that some children are better able to memorise many words than to learn the logic of segmenting and sequencing and conversely others may find the logical approach easier then memorising many words? If this is the case, it could indicate that two teaching methods are required based on the child's way of thinking. This is a complex question based only on intuition but could be worthy of further investigation.

References

References

ADAMS, M. 1990. *Beginning to Read: Thinking and Learning about Print*, Cambridge, MA: MIT Press.

ALS, B., JENSEN, J. & SKOV, M. 2005. Comparison of Think-Aloud and Constructive Interaction in Usability Testing with Children. *Interaction Design and Children 2005 conference.* Boulder, Colorado ACM.

AUGUR, J. & BRIGGS, S. 1992. Hickey Multi-Sensory Language Course. London: Whurr Publishers Ltd.

BALOGH, J. 2001. Strategies for concatenating recordings in a voice user interface: what we can learn from prosody. . *CHI '01 extended abstracts on Human factors in computing systems,.* Seattle, Washington,.

BEARD, R. 1999. *National Literacy Strategy, Review of Research and other Related Evidence* [Online]. Available: http://www.standards.dfes.gov.uk/primary/publications/literacy/63541/nls_reviewresearch_nlsrr.pdf [Accessed 10/12/10].

BECKER, L. 2000. *Effect Size (ES)* [Online]. Available: http://www.uccs.edu/~faculty/lbecker/es.htm [Accessed 24/12/2010].

BLAKEMORE, D., WINSTON, J. & FRITH, U. 2004. Social cognitive neuroscience: where are we heading? . *Trends in Cognitive Sciences,* 8**,** 216-222.

BORUCH, R., DE MOYA, D. & SNYDER, B. 2002. The Importance of Randomized Field Trials in Education and Related Areas. *In:* MOSTELLER, F. & BORUCH, R. (eds.) *Evidence Matters: Randomized Trials in Education Research.* Washington: Brookings Institution Press.

BRINDLEY, S. 2000. ICT and Literacy. *In:* GAMBLE, N. & EASINGWOOD, N. (eds.) *ICT and Literacy: Information and Communications Technology, Media, Reading and Writing.* London: Continuum.

BROOKS, G. 2002a. Phonemic awareness is a key factor in learning to be literate: how best should it be taught? *In:* COOK, M. (ed.) *Perspectives on the Teaching and Learning of Phonics.* Royston, Herts: UK Reading Association.

BROOKS, G. 2002b. *What Works for Children with Literacy Difficulties? The Effectiveness of Intervention Schemes* [Online]. Available: http://www.dfes.gov.uk/research/data/uploadfiles/RR380.pdf [Accessed 15/12/10].

BROOKS, G. 2003. Sound Sense: the phonics element of the National Literacy Strategy. A report to the Department for Education and Skills. . London:DfES.

BROOKS, G. 2007a. What works for pupils with literacy difficulties? The effectiveness of intervention schemes. *In:* DCSF (ed.) Third ed.

BROOKS, G. 2007b. *What Works for Pupils with Literacy Difficulties? The Effectiveness of Intervention Schemes. 3rd Edition* [Online]. Available: http://www.standards.dfes.gov.uk/primary/publications/literacy/pri_lit_what_works0068807/pri_lit_what_works0068807.pdf [Accessed 10/12/10].

BROOKS, G., FLANAGAN, N., HENKHUZENS, Z. & HUTCHISON, D. 1998. *What Works for Slow Readers? The Effectiveness of Early Intervention Schemes.* [Online]. Available: http://www.nfer.ac.uk/publications/pdfs/downloadable/Whatworksforslowreaders.pdf [Accessed 12/12/10].

BROOKS, G., MILES, J., TORGERSON, C. J. & TORGERSON, D. J. 2006. Is an intervention using computer software effective in literacy learning? A randomised controlled trial, . *Educational Studies.*, 32**,** 133-143.

CHIASSON, S. & GUTWIN, C. 2005. Testing the Media Equation with Children *SIGCHI conference on Human factors in computing systems.* Portland, Oregon, USA

CLARK, C. & AKERMAN, R. 2008. *Being a Reader: The relationship with gender* [Online]. National Literacy Trust. Available: http://www.literacytrust.org.uk/assets/0000/0567/Being_reader_gender_2008.pdf [Accessed 01/10/10].

CLISSETT, P. 2008. Evaluating qualitative research. *Journal of Orthopaedic Nursing,* 12**,** 99-105.

COHEN, J. 1992. A power primer. *Psychological Bulletin,* 112**,** 155-159.

CONDIE, R., MUNRO, B., SEAGRAVES, L. & KENESSON, S. 2007. *The impact of ICT in schools - a landscape review,* [Online]. Quality in Education Centre, University of Strathclyde. Available: http://publications.becta.org.uk/download.cfm?resID=28221 [Accessed 12/11/10].

CONSTANTINE, L. 2003. CHI 2003 Feature: Testing. 1 2 3 4 5. Testing. .

CORDES, C. & MILLER, E. 2000. *Fool's Gold: A Critical Look at Computers in Childhood. Alliance for Childhood* [Online]. Available: http://www.eric.ed.gov/PDFS/ED445803.pdf [Accessed 05/01/11].

COX, M., WEBB, M., ABBOTT, C., BLAKELEY, B., BEAUCHAMP, T. & RHODES, V. 2004. A review of the research literature. *ICT and Pedagogy.*: DfES.

CRAVEN, R., MARSH, H., JAYASINGHE, U. & DEBUS, R. 2001. Diffusion effects : Control group contamination threats to the validity of teacher-administered interventions. *Journal of educational psychology* 93**,** 639-645.

CUBAN, L. 2001. Oversold and Underused: Computers in the Classroom. Harvard University Press, Cambridge, MA.

DARVES, C., OVIATT, S. & COULSTON, R. 2002. The Impact of Auditory Embodiment on Animated Character Design. *Autonomous Agents and Multiagent Systems (AAMAS '02).* Bologna, Italy.

DAVISON, G., MURPHY, S. & WONG, R. 2005. The use of eBooks and interactive multimedia as alternative forms of technical documentation. *23rd annual international conference on Design of communication.* Coventry, United Kingdom: ACM.

DEL BOCA, F. K. & DARKES, J. 2007. Enhancing the validity and utility of randomized clinical trials in addictions treatment research: III. Data processing and statistical analysis. *Addiction,* 102**,** 1356-1364.

DFES. 2007a. *Criteria for assuring high quality phonic work* [Online]. Available: http://www.standards.dfes.gov.uk/phonics/programmes/core/ [Accessed 01/11/2010].

DFES. 2007b. *Letters and Sounds: Principles and Practice of High Quality Phonics Notes of Guidance for Practitioners and Teachers* [Online]. Available: http://nationalstrategies.standards.dcsf.gov.uk/node/84969 [Accessed 15/12/10].

DFES. 2008. *Phonics Programme Publisher self-assessment template* [Online]. Available: http://www.education.gov.uk/schools/teachingandlearning/pedagogy/a0010231/choosing-an-effective-phonics-teaching-programme [Accessed 20/12/10].

DFES. 2010. *Gender and achievement resources* [Online]. Available: http://nationalstrategies.standards.dcsf.gov.uk/node/46125 [Accessed 01/11/2010].

DIX, A., FINLAY, J., ABOWD, G. & BEALE, R. 2004 *Human-Computer Interaction*, Prentice Hall.

DONKER, A. & REITSMA, P. 2004. Usability Testing With Young Children. *Interaction Design and Children.* Maryland, USA.

DRAKE, L. & HEPPLEWHITE, D. 2004. Take the letter shapes out of the water and the word cards out of the sand - Review of ‘Playing with Sounds’: a supplement to ‘Progression in Phonics’. . *Reading Reform Foundation Newsletter, no.53 (Autumn).*

EARL, L., FULLAN, M., LEITHWOOD, K. & WATSON, N. 2000. *Watching and learning: OISE/UT evaluation of implementation of the National Literacy and Numeracy Strategies* [Online]. [Accessed 12/12/10].

EDUCATION AND SKILLS COMMITTEE HOUSE OF COMMONS. 2005. *‘Teaching Children to Read’ Eighth Report of Session 2004-05* [Online]. Available: Published on the Parliament Publications and Records Site http://www.publications.parliament.uk/pa/cm200405/cmselect/cmeduski/121/121.pdf [Accessed 20/12/10].

EDUCATION AND SKILLS SELECT COMMITTEE. 2005. *First Special Report* [Online]. Available: http://www.publications.parliament.uk/pa/cm200506/cmselect/cmeduski/406/40605.htm [Accessed 10/12/10].

EDWARDS, H. & BENEDYK, R. 2007. A Comparison of Usability Evaluation Methods for Child Participants in a School Setting. *Interaction Design and Children 6th international conference.* Aalborg, Denmark

EHRI, L. C., NUNES, S. R., STAHL, A. S. & WILLOWS, D. A. 2001. Systematic phonics instruction helps children learn to read: evidence from the National Reading Panel's meta-analysis. *Review of Educational Research,* 71**,** 393-447.

ELDRIDGE, S. M., ASHBY, D. & KERRY, S. 2006. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *International Journal Of Epidemiology,* 35**,** 1292-1300.

ELLIS, S. 2007. Policy and research: Lessons from the Clackmannanshire phonics initiative. *Journal of Early Childhood Literacy,* 7**,** 281-297.

EPPI. 2010. *The Evidence for Policy and Practice Information and Co-ordinating Centre* [Online]. Available: http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=53&language=en-US [Accessed 01/10/2010].

EVANGELOU, M., BROOKS, G., SMITH, S. & JENNINGS, D. 2005. *The Birth to School Study: A Longitudinal Evaluation of the Peers Early Education Partnership (PEEP) 1998-2005* [Online]. Available: http://www.peep.org.uk/standard.asp?id=1145 [Accessed 01/11/2010].

FAUL, F., BUCHNER, A., ERDFELDER, E. & LANG, A.-G. 2009. *G*Power* [Online]. Available: http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/ [Accessed 10/12/10].

FAUL, F., ERDFELDER, E., LANG, A.-G. & BUCHNER, A. 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods,* 32**,** 175-191.

FAULKNER, L. 2003. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behaviour Research Methods, Instruments & Computers,* 35**,** 379 - 383.

FEWTRELL, M. S., KENNEDY, K., SINGHAL, A., MARTIN, R. M., NESS, A., HADDERS-ALGRA, M., KOLETZKO, B. & LUCAS, A. 2008. How much loss to follow-up is acceptable in long-term randomised trials and prospective studies? *Archives of Disease in Childhood,* 93**,** 458-461.

FUJII, K., OKAWA, J. & SUIGETSU, K. 2007. High-Individuality Voice Conversion Based on Concatenative Speech Synthesis. *Proceedings of World Academy of Science: Engineering & Technology,* 36**,** 483-488.

GILBERT, C. 2007. *The Gilbert Review* [Online]. DfES Publications. Available: http://publications.teachernet.gov.uk/eOrderingDownload/6856-DfES-Teaching%20and%20Learning.pdf [Accessed 20/12/10].

GILUTZ, S. & NIELSEN, J. 2002. *Usability of Websites for Children: 70 Design Guidelines* [Online]. Available: Published at: http://www.nngroup.com/reports/kids/ [Accessed 20/12/10].

GRAY, J. 2003. What Next? A Dozen Information-Technology Research Goals. *Journal of the ACM,* 50**,** 41-57.

HANNA, L., RISDEN, K. & ALEXANDER, K. 1997. Guidelines for usability testing with children, . *Interactions, (September+October).*

HAUGLAND, S. W. 1999. What role should technology play in young children's learning? . *Young Children,* 54**,** 26-31.

HEALY, J. 1998. Failure To Connect: How Computers Affect Our Children's Minds--for Better and Worse. *Computers and Education.* Simon & Schuster, New York.

HIGGINS, S. 2003. *Does ICT Improve Learning and Teaching in Schools? A Professional User Review for BERA*, Notts. BERA.

HOURCADE, J., BEDERSON, B. & DRUIN, A. 2004. Preschool Children's Use of Mouse Buttons. *CHI 2004: Extended Abstracts of Human Factors in Computing Systems (CHI 2004).* Vienna, Austria ACM Press.

HOURCADE, J., CROWTHER, M. & HUNT, L. 2007. Does Mouse Size Affect Study and Evaluation Results? A Study Comparing Preschool Children's Performance with Small and Regular-Sized Mice, . *Interaction Design and Children 6th international conference.* . Aalborg, Denmark.

HUTCHISON, D. & STYLES, B. 2010. *A Guide to Running Randomised Controlled Trials for Educational Researchers.* [Online]. Slough: NFER. Available: http://www.nfer.ac.uk/publications/RCT01/ [Accessed 5/01/11].

IPA. 2005. *International Phonetic Association* [Online]. Available: http://www.langsci.ucl.ac.uk/ipa/ [Accessed 01/11/2010].

JACOBY, L. 1978. On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behaviour,,* 17**,** pp 649 - 667.

JENSON, J. & SKOV, M. 2005. A Review of Research Methods in Children's Technology Design. *Interaction Design and Children 2005 conference* Boulder, Colorado: ACM.

JOHNSTON, R. & WATSON, J. 2005. *The effects of synthetic phonics teaching on reading and spelling attainment: a seven year longitudinal study* [Online]. Available: http://www.scotland.gov.uk/Resource/Doc/36496/0023582.pdf [Accessed 20/12/10].

JOHNSTON, R. S. & WATSON, J. E. 2004. Accelerating the development of reading, spelling and phonemic awareness skills in initial readers. *Reading & Writing,* 17**,** 327-357.

KÄHKÖNEN, M. & OVASKA , S. 2006. Initial Observations on Children and Online Instructions. *Interaction Design and Children 2006 conference.* Tampere, Finland

KEHOE, A. & PITT, I. 2006. Designing Help Topics for use with Text-To-Speech,. *24th annual ACM international conference on Design of communication* Myrtle Beach, USA ACM.

KENNEDY, S. 1989. Using video in the BNR usability lab *ACM SIGCHI Bulletin,* 21**,** 92-95.

KILLIP, S., MAHFOUD, Z. & PEARCE, K. 2004. What Is an Intracluster Correlation Coefficient? Crucial Concepts for Primary Care Researchers. *Ann Fam Med,* 2**,** 204-208.

KINDER, M. 1991. *Playing with Power in Movies, Television and Video Games: From Muppet Babies to Teenage Mutant Ninja Turtles.* , University of California Press, Berkeley, CA.

KITCHEN, S., FINCH, S. & SINCLAIR, R. 2007. *Harnessing Technology schools survey* [Online]. British education communication technology agency (BECTA). Available: http://schools.becta.org.uk/upload-dir/downloads/page_documents/research/harnessing_technology_schools_survey07.pdf [Accessed 20/12/10].

KITCHEN, S., MACKENZIE, H., BUTT, S. & FINCH, S. 2006. *Evaluation of Curriculum Online Report of the third survey of schools* [Online]. British education communication technology agency (BECTA). Available: http://research.becta.org.uk/upload-dir/downloads/page_documents/curriculum_online_third_survey0406.pdf [Accessed 20/11/10].

KOPPITZ, E. M. 1970. The visual aural digit span test with elementary school children. *Journal of Clinical Psychology,* 26**,** 349-353.

LANKSHEAR, C. & KNOBEL, M. 2003. New technologies in early childhood literacy research: a review of the research. *Journal of Early Childhood Literacy,* 31**,** 59-82.

LEON, A. C., MALLINCKRODT, C. H., CHUANG-STEIN, C., ARCHIBALD, D. G., ARCHER, G. E. & CHARTIER, K. 2006. Attrition in Randomized Controlled Clinical Trials: Methodological Issues in Psychopharmacology. *Biological Psychiatry,* 59**,** 1001-1005.

LEVENTHAL, L. & BARNES, J. 2007. *Usability Engineering: Process, Products and Examples*, Pearson Prentice Hall.

LI, G. & LAI, J. 2001. Shall We Mix Synthetic Speech and Human Speech? Impact on Users' Performance, Perception, and Attitude. *CHI '01*. Seattle, Washington.

MACFARLANE, S., SIM, G. & HORTON, M. 2005. Assessing Usability and Fun in Educational Software. *Interaction Design and Children.* Boulder, Colorado: ACM.

MARKOPOULOS, P., READ, J., MACFARLANE, S. & HOYSNIEMI, J. 2008. *Evaluating Children's Interactive Products: Principles and Practices for Interaction Designers*, Morgan Kaufmann.

MITCHELL, M. & FOX, B. 2001. The effects of computer software for developing phonological awareness in low-progress readers. *Reading Research and Instruction*, 40**,** 315-332.

MIYAKE, N. 1986. Constructive Interaction and the Iterative Process of Understanding, . *Cognitive Science,* 10**,** 151-177.

NATIONAL READING PANEL. 2000. *Teaching Children to Read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the sub-groups* [Online]. Available: http://www.nichd.nih.gov/publications/nrp/report.cfm [Accessed 20/11/10].

NICOL, A. & CASEY, C. 2003. CAL Interface Design for Young Children: a Case Study in Literacy. *CAL '03.* Queen's University, Belfast: Elsevier Science.

NICOL, A. & SNAPE, L. 2003. A computer based letter formation system for children. *Interaction Design and Children 2003.* Preston, UK.

NICOL, A. & SNAPE, L. 2007. Interface Design Guidelines for Pre-Reading Children. *CAL '07.* Trinity College, Dublin: Elsevier Science.

NIELSEN, J. 1993. *Usability engineering* Academic Press, Boston.

NIELSEN, J., CLEMMENSEN, T. & YSSING, C. 2002. Getting access to what goes on in people's heads? reflections on the think-aloud technique. *2nd Nordic conference on Human-computer interaction.* Aarhus, Denmark: ACM.

NIEMI, H. & OVASKA, S. 2007. Designing Spoken Instructions with Preschool Children *Interaction Design and Children 6th international conference* Aalborg, Denmark: ACM.

NORMAN, D. 1988 *The Psychology of Everyday Things*, Perseus Publishing.

NUT REPORT. 2009. *NUT report: special measures schools and deprivation - March 2009* [Online]. National Union of Teachers. Available: http://www.teachers.org.uk/node/11417 [Accessed 05/01/11].

NUTBROWN, C. & HANNON, P. 1993. Assessing early literacy -new methods needed. *International Journal of Early Childhood.,* 25**,** 27-30.

NUTBROWN, C., HANNON, P. & MORGAN, A. 2005. *Early Literacy Work with Families*, Sage Publications.

OFSTED 1998. *The National Literacy Project: An HMI Evaluation*, London: Ofsted.

OFSTED. 2005. *Embedding ICT in schools - a dual evaluation exercise* [Online]. Available: http://www.ofsted.gov.uk/Ofsted-home/Publications-and-research/Browse-all-by/Education/Curriculum/Information-and-communication-technology/Secondary/Embedding-ICT-in-schools-a-dual-evaluation-exercise [Accessed 20/12/10].

OFSTED. 2006. *The initial training of further education teachers - Findings from 2004/05 inspections.* [Online]. Available: http://www.ofsted.gov.uk/publications/index.cfm?fuseaction=pubs.summary&id=4153 [Accessed 20/11/10].

OFSTED. 2002. *The National Literacy Strategy: the first four years 1998-2002.* [Online]. Available: http://www.standards.dfes.gov.uk/primary/publications/literacy/63649/hmi555_nls_first4years.pdf [Accessed 20/11/10].

PAPERT, S. 2004. The challenges of IDC: What have we learned from our past? A conversation with Seymour Papert, Marvin Minsky and Alan Kay. *Communications of the ACM January 2005* 48.

PLOWMAN, L. & STEPHEN, C. 2003. A 'benign addition'? Research on ICT and pre-school children. *Computer Assisted Learning*, 19**,** 149-164.

POLIT, D. F. & GILLESPIE, B. M. 2010. Intention-to-treat in randomized controlled trials: Recommendations for a total trial strategy. *Research in Nursing & Health*, 33**,** 355-368.

PREECE, J., ROGERS, Y., SHARP, H., HOLLAND, S. & CAREY, T. 1994 *Human-Computer Interaction*, Addison-Wesley.

RAPEEPISARN, K., WONG, K., FUNG, C. & DEPICKERE, A. 2006. Similarities and differences between "learn through play" and "edutainment". *3rd Australasian conference on Interactive entertainment.* Perth, Australia: ACM.

READ, J. & MACFARLANE, S. 2006. Using the Fun Toolkit and Other Survey Methods to Gather Opinions in Child Computer Interaction. *Interaction Design and Children* Tampere, Finland: ACM.

ROSE, J. 2006. *Independent Review of the Teaching of Early Reading. Final Report.* [Online]. Available: www.standards.dfes.gov.uk/rosereview [Accessed 20/11/10].

ROUSE, C. & KRUEGER, A. 2004. *Putting computerized instruction to the test: a randomized evaluation of a "scientifically based" reading program. Working paper.* [Online]. Elsevier. Available: http://www.ers.princeton.edu/workingpapers/5ers.pdf [Accessed 05/11/2010].

SCHMIDER, E., ZIEGLER, M., DANAY, E., BEYER, L. & BÜHNER, M. 2010. Is It Really Robust?: Reinvestigating the Robustness of ANOVA Against Violations of the Normal Distribution Assumption. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences,* 6**,** 147-151.

SCHULZ, K. F., ALTMAN, D. G. & MOHER, D. 2010. CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials. *PLoS Clinical Trials,* 6**,** 1-7.

SHARE, D. L. 2004. Knowing letter names and learning letter sounds: A causal connection. *Journal of Experimental Child Psychology.,* 88.**,** 213-233.

SHENTON, A. & PAGETT, L. 2007. From 'bored to screen': the use of the interactive whiteboard for literacy in six primary classrooms in England. *Literacy,* 41**,** 129-136.

SLAVIN, R., LAKE, C., CHAMBERS, B., CHEUNG, A. & DAVIS, S. 2009. *Effective Beginning Reading Programs: A Best-Evidence Synthesis* [Online]. Available: http://www.bestevidence.org/reading/begin_read/keyfind.htm [Accessed 01/11/2010].

SNAPE, L. 2007. Phonics for Young Children: A Computer Based Approach. *CAL '07.* Trinity College, Dublin.

SNAPE, L. & CASEY, C. 1997. Using Speech in Multimedia Applications. *TCD Conference on Multimedia.*

SNAPE, L. & NICOL, A. 2003a. Evaluating the effectiveness of a computer-based letter formation system for children. *Interaction Design and Children.* Preston, UK: ACM.

SNAPE, L. & NICOL, A. 2003b. Using CAL to Aid in the Development of Letter Formation Skills in Young Children. *CAL '03.* Queen's University, Belfast: Elsevier Science.

SNAPE, L., NICOL, T. & MACFARLANE, S. 2003. Phonics for Young Children: a Computer Based Approach. *Interaction Design and Children 2003.* Preston, UK: ACM.

SNOWLING, M., STOTHARD, S., CLARKE, P., BOWYER-CRANE, C., HARRINGTON, A., TRUELOVE, E. & HULME, Y. 2009. *York Assessment of Reading for Comprehension (YARC),* [Online]. G L Assessment,. Available: http://www.yarcsupport.com/about.html [Accessed 20/10/2010].

STECKLER, A., MCLEROY, K., R., GOODMAN, R., M., BIRD, S., T. & MCCORMICK, L. 1992 Toward integrating qualitative and quantitative methods: an introduction. . *Health Education Quarterly,* 19**,** 1-8.

THRASS. 2009. *Phoneme Machine 6.1* [Online]. THRASS (UK) LTD. Available: http://www.phonememachine.com/ [Accessed 05/06/10].

TIAN-SWEE, T. 2009. Corpus Design for Malay Corpus-based Speech Synthesis System. *American Journal of Applied Sciences,* 6**,** 696-702.

TORGERSON, C. & TORGERSON, B. 2007. The Need for Pragmatic Experimentation in Educational Research. *Econ. Innov. New Techn.,* 16**,** 323-330.

TORGERSON, C. & TORGERSON, D. 2003. Avoiding Bias in Randomised Controlled Trials in Educational Research. *British Journal of Educational Studies,* 51**,** 36-45.

TORGERSON, C. & ZHU, D. 2003. *A systematic review and meta-analysis of the effectiveness of ICT on literacy learning in English, 5-16.* [Online]. Available: http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=198 [Accessed 5/11/2010].

TORGERSON, C. J. 2009. Randomised controlled trials in education research: a case study of an individually randomised pragmatic trial. *Education 3-13: International Journal of Primary, Elementary and Early Years Education,* 37**,** 313-321.

TORGERSON, C. J., BROOKS, G. & HALL, J. 2006. *A Systematic Review of the Research Literature on the Use of Phonics in the Teaching of Reading and Spelling - Full report.* [Online]. Available: http://nationalstrategies.standards.dcsf.gov.uk/node/84443?uc=force_uj [Accessed 05/01/11].

TORGERSON, C. J. & ELBOURNE, D. 2002. A systematic review and meta-analysis of the effectiveness of information and communication technology (ICT) on the teaching of spelling. *Journal of Research in Reading,* 25**,** 129.

TORGERSON, D. 2007. *What makes a good quality trial?* [Online]. Available: http://www.restore.ac.uk/trials-pp/Workshops/David%20Torgerson.ppt [Accessed 24/12/2010].

UNDERWOOD, J. 1994. Where are we now and where are we going? *Computer Based Learning: Potential into Practice.* David Fulton Publishers.

UNDERWOOD, J. 2000. A Comparison of Two Types of Computer Support for Reading Development. *Journal of Research in Reading,* 23**,** 136-148.

UNDERWOOD, J., BAGULEY, T., BANYARD, P., COYNE, E., FLINT, L. & SELWOOD, I. 2007. *Impact 2007: Personalised Learning with Technology* [Online]. Available: http://www.Becta.org.uk [Accessed 20/11/10].

UNDERWOOD, M., BARNETT, A. & HAJIOFF, S. 1998. Cluster randomization: a trap for the unwary. *Br J Gen Pract.,* 48**,** 1089-1090.

VALENTINE, G., MARSH, J. & PATTIE, C. 2005. *Children and Young People's Home Use of ICT for Educational Purposes: The Impact on Attainment at Key Stages 1-4* [Online]. Available: http://www.dcsf.gov.uk/research/data/uploadfiles/RR672.pdf [Accessed 20/11/10].

VAN KESTEREN, I., BEKKER, M., VERMEEREN, A. & LLOYD, P. 2003. Assessing Usability Evaluation Methods on their Effectiveness to Elicit Verbal Comments from Children Subjects. *Interaction Design and Children.* Preston, UK.: ACM.

VAN SANTEN, J., BLACK, L., COHEN, G., KAIN, A., KLABBERS, E., MISHRA, T., DE VILLIERS, J. & NIU, X. 2003. Applications of Computer Generated Expressive Speech for Communication Disorders. *Eurospeech 2003.* Geneva.

VAN SCOTER, J., ELLIS, D. & RAILSBACK, J. 2001. *Technology in Early Childhood Education:finding the balance* [Online]. Available: http://www.netc.org/earlyconnections/byrequest.pdf [Accessed 20/12/10].

VOLMAN, M., ECK, E., HEEMSKERK, I. & KUIPER, E. 2005. New technologies, new differences. Gender and ethnic differences in pupils' use of ICT in primary and secondary education. *Computers & Education,* 45**,** 35 - 55.

WALLER, T. 2006. *Literacy and ICT in the Early Years in ICT in the Early Years,* McGraw Hill.

WILD, M. 2009. Using computer-aided Instruction to support the systematic practise of phonological skills in beginning readers *Journal of Research in Reading,* 32**,** 413-432.

WOOD, C. 2005. Beginning readers' use of talking books software can affect their reading strategies. *Journal of Research in Reading,* 28**,** 170-182.

ZAMAN, B. & ABEELE, V. 2007. Towards a Likeability Framework that meets Child-Computer Interaction & Communication Sciences. *Interaction Design and Children 6th international conference.* Aalborg, Denmark: ACM.

Appendix A

Child-Computer Interaction

Appendix A Child-Computer Interaction

A1 Introduction

The literature review detailed in Chapter 2 supports the premise that a computer application to teach phonemic awareness could be a useful classroom tool. An experiment to determine the effectiveness of such an application is the subject of Chapter 6. To rule out poor usability as an uncontrolled variable it was necessary to develop and evaluate an effective interface for a phonemic awareness application to be used in the randomised controlled trial. Computer applications that are difficult to use will also place an additional burden on the teaching workload if teachers are required to support the computer sessions. It is also important that children engage with computer applications and enjoy using them.

Child-Computer Interaction (ChiCI) is a relatively new and developing field of research. Hanna et al (1997) reported on the area of Child-Computer Interaction that *"the usability of a product is closely related to children's enjoyment of it."*

Work in the area of designing "enjoyable" and "fun" applications for children has continued (MacFarlane et al., 2005, Rapeepisarn et al., 2006, Zaman and Abeele, 2007). MacFarlane et al (2005) concluded from their study of children using computer applications that usability is a pre-requisite in the development of engaging and fun applications:

> *"Our observations showed that the children appeared to have less fun when their interactions had more usability problems."*

There was little research in the area of Human Computer Interaction for young children when this phase of the project was started and it is still a developing area (Markopoulos et al., 2008). To address this it was necessary to carry out research into child-computer interaction with particular emphasis on the usability of event-driven graphical user interface applications, the interaction mode of modern systems. The product of that research is a set of design guidelines which can be used to aid the development of desktop multimedia-based computer software for young children. This appendix discusses the development of the interaction design guidelines.

Gilutz and Nielsen (2002) defined a set of interaction guidelines based on findings from a usability study of children's websites for older children. However, the youngest children involved in their tests were between six and eight years, which would mean they would mostly have been more advanced readers than the children in this study, who were aged between 5 and 6 years. In the absence of child-computer interaction guidelines for children younger than 5 years of age, it was necessary to research and develop a set on which to base the design for the phonemic awareness computer application developed for this project with a target age group of children aged 5 to 6 years.

A2 Design Guidelines Development

This appendix discusses the difficulties of designing computer interfaces for young children - particularly beginner-readers. Widely published methodologies and guidelines for **adult** Human Computer Interaction (HCI), (Norman, 1988 , Preece et al., 1994 , Dix et al., 2004) can be applied to the development of interfaces for young children but there are some areas that need to be dealt with differently. For example, pre-reading aged children can't read instructions.

Iterative design was used throughout the development of the applications discussed in this appendix:

- Computerised educational activities were developed based on analysis of users and age-appropriate tasks were chosen.
- The software was fault tested using established software testing methods.
- Defects found during this stage were fixed to ensure that software errors would not affect the usability evaluations.
- Activities were tested for usability on children of an appropriate age group and findings were recorded.
- Activities were refined according to the findings and re-tested using the revised ideas (Nicol and Snape, 2003, Snape and Nicol, 2003b, Snape and Nicol, 2003a, Nicol and Casey, 2003).

The iterative refinement method used for these applications resulted in a set of published child-computer interaction guidelines (Nicol and Snape, 2007). The guidelines originated from an initial set created through discussions with teachers from several primary schools (who understood the needs of the age of child that they teach) and interviews with developers of children's educational applications (who understand what can be done technically).

A prototype application, "Letterworld", using hypothesised guidelines to test them was initially developed; an example of two of the Letterworld screens is provided in Figure A-1. Discussion with primary school teachers at the outset of the guideline development identified a need for a letter formation application; the required one-to-one interaction between instructor and tutor was very time consuming so a computer application could be a useful teaching aid.

The developed application requires the child to select a letter of the alphabet from the main screen then click the segments in the letter in the correct sequence. When complete, the child is asked to find objects in the picture which begin with that letter; this mainly adds a fun element to the application as a reward for completing the letter correctly.

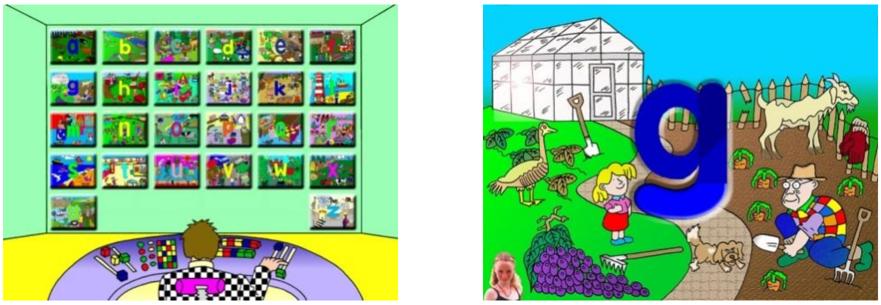

Figure A-1: Letterworld home screen and one of the letter formation activities

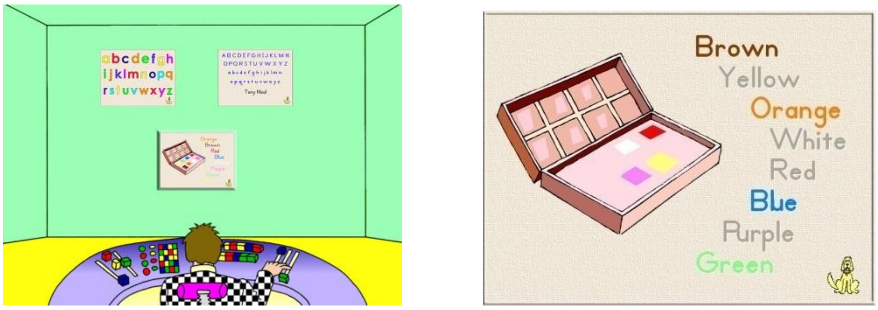

Figure A-2: Literacyworld home screen and one of the activity screens

Figure A-3: Vocabulary Tutor activity screen

The guidelines were updated as a result of the evaluation of the Letterworld screens and re-evaluated on three more applications. In total, four applications were developed using the guidelines: Letterworld (Figure A-1) developed by Linda Snape, Literacyworld (Figure A-2) and a Vocabulary tutor using speech recognition (Figure A-3) developed by Tony Nicol and the Phonics tutor developed by Linda Snape for the RCT in this project (Figure A-4). The development of each application enabled refinement of the interaction guidelines set.

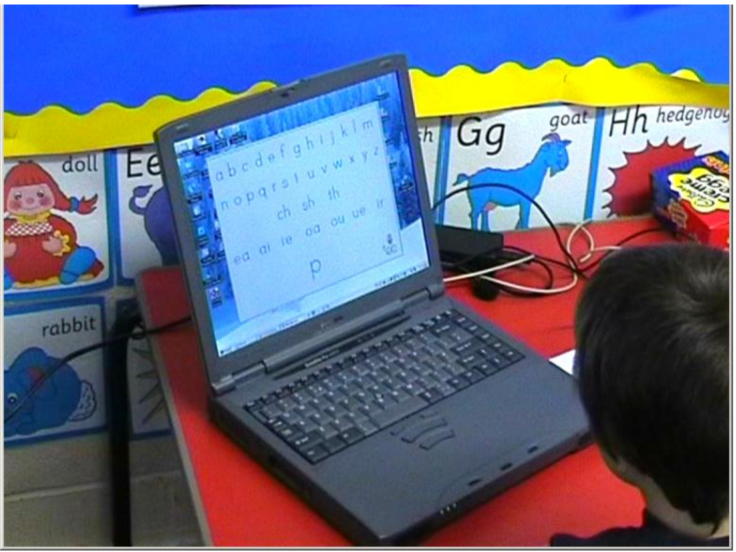

Figure A-4: SPEL session during evaluation

The following child-computer interaction design guidelines were developed through the iterative design process used during the evaluation of the prototype applications:

- Minimise the use of text as a feedback mechanism.
- Minimise the number of interactive controls.
- Vary the feedback.
- Minimise keyboard input.
- Make use of randomness.
- Enable early exit from the application which is inaccessible to children.
- Provide tiered pro-active interactive help.
- Enable user-defined options but make them inaccessible to children.
- Ensure that interaction hardware is of an appropriate size for the user.
- Design for no scrolling.

Guidelines that are general to the field of HCI are not discussed, as standard HCI principles are well understood – the guidelines listed are additions to HCI to cater specifically for young children. The developed design guidelines are listed in the following sections and are compared where appropriate to guidelines proposed by other researchers.

A2.1 Minimise the use of text as a feedback mechanism

Designing interfaces for non-readers or beginner-readers necessitates a shift in design principles for older users who can read. Text based instructions, text based feedback and controls such as buttons requiring textual descriptions are of no use for beginning readers. Kähkönen and Ovaska (2006) reported on usability evaluations of computer applications with young children, finding that even for children who could read, written instructions still posed problems. The use of graphical feedback is appropriate to this age group if the icons clearly convey meaning.

Spoken output provides an alternative or supplementary approach for beginner-readers and the two choices of spoken output on a computer interface are text-to-speech or digital samples of human voices. Gray (2003) from Microsoft Research, cites text-to-speech on his list of *"Dozen long-term system research problems"*, proposing that researchers must strive to make computers *"Speak as well as a native speaker"*. Darves et al (2002) found that text-to-speech (TTS) gave positive results when used for output with children as young as 7. However, the researchers reported that the teachers involved with this project felt that it was unsuitable for children of this age and younger; a sentiment that was echoed by the primary school teachers interviewed for this research project. Studies carried out by Li and Lai (2001) reported that even adult users have difficulty in understanding text-to-speech output. In terms of software development, providing synthesised speech as part of the user interface is easier to implement than providing human speech (Snape and Casey, 1997). However, both published research and discussions with teachers during this research project suggested that text-to-speech is not the preferred option for young children as it is too robotic and is considered particularly unsuitable for systems that rely on correct pronunciation, such as literacy and phonics tutors. As Van Santen et al (2003) point out "*Generating meaningful and natural sounding prosody is a central challenge in text-to-speech synthesis (TTS)*". Although text-to-speech quality has improved over the last few years (Tian-Swee, 2009), the use of human speech was found, during the evaluation sessions of this study, to be the most suitable solution for a child-computer interface and children showed a preference for the non-dominant voice of a child to that of an adult. A similar observation was reported by Darves et al (2002). Current realistic-sounding computer generated speech uses a concatenation of human speech samples (Fujii et al., 2007) which is a similar but more granular approach to that used by SPEL.

Gilutz and Nielsen (2002) explain that a few seconds of clearly recorded audio in age-appropriate language should be used for non-readers. However, they warn against using audio clips for navigation on web-sites, as download delays of audio clips may mean that a child may have made a choice before the audio has been downloaded and played if the download is delayed. Similarly, Kähkönen and Ovaska (2006) state:

> *"while not feasible on the web with limited bandwidth connections, audio help seems to overcome the limitations of written instructions."*

Although the risk of delay is much lower in standalone applications than in web applications, all controls developed as part of this study have spoken and graphical feedback.

Gilutz and Nielsen (2002) reported that children became irritated as the long audio clips used to provide explicit directions slowed them down. Similar results were found in this research project when evaluating "Letterworld" which was the first prototype application. Two possible solutions to this problem were implemented and evaluated: provide a facility that is accessible to the teacher to disable lengthy explanations for familiar users, and provide a "barge-in[3]" feature for familiar users. The facility to disable long explanations can be used to set the instructional audio to be played each time the programme is run, only the first time it is run or never to play. The "barge-in" feature will terminate the audio if the user presses the next control (with the mouse or the keyboard); this gives the user the control to listen to the instructions only if they need to.

A2.2 Minimise the number of interactive controls

Young children like to experiment and this should be encouraged as long as designers are aware that some children will click around the screen. To provide a simple and

[3] Barge-in is where the speech is cancelled if the user continues to interact with the application

intuitive interface, the only accessible controls should be those which are part of the activity and at most a single help control.

Although visibility and affordance are standard Graphical User Interface (GUI) elements of HCI design for adults or children (Norman, 1988 , Dix et al., 2004), there may be different implementation requirements for interfaces designed for young children, mainly because text is not an acceptable form of communication. It is important that each control provides either verbal or unambiguous graphical feedback to the user. For example, if the mouse rolls over a control, the control could brighten to indicate it is active and where necessary, speech could be output to indicate the purpose of the of the control.

Icons and graphics in all of applications developed during this study were designed to be intuitive and natural and wherever possible the icons use a one-to-one mapping. The letter thumbnails used on the home screen of Letterworld, illustrated in Figure A-5, provide a good example of visibility and affordance: inactive buttons appear greyed out; active letters are three-dimensional, and when rolled over, become brighter; the shape of the cursor changes to a pointed finger as the buttons are rolled over to give visual feedback to the user.

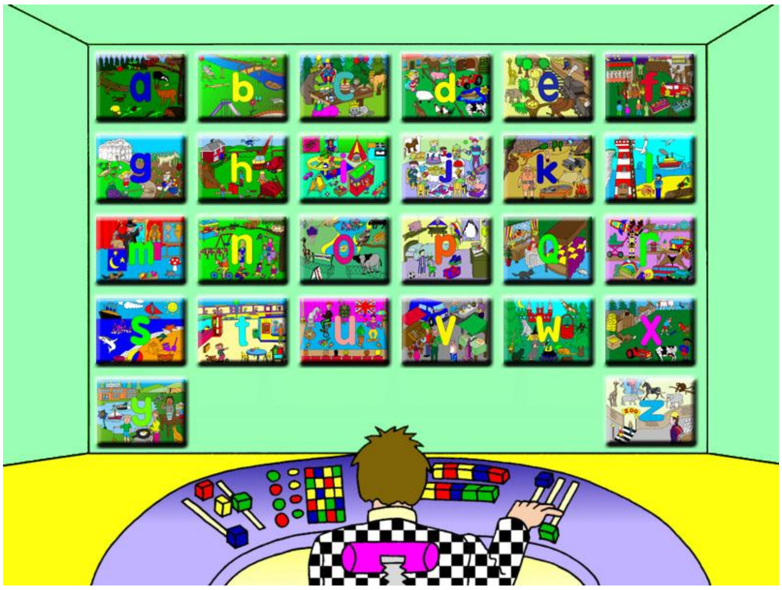

Figure A-5: Alphabet Activity Screen

It is important that GUIs provide users with clear controls with which to navigate smoothly around the application. In the case of interfaces for young children, these controls need to be labelled with graphics and/or voice to ensure navigation is natural and intuitive. Navigation controls provided on the main menu screen of Letterworld (illustrated in Figure A-5) take the form of buttons displaying a small graphical representation of the activity screen to which they will navigate. The controls also provide verbal feedback relating to their function by sounding out the letter when rolled over and other verbal navigational clues, for example, sounding out phrases such as "here we go" when moving into an activity.

Gilutz and Nielsen (2002) claim that users should be able to find what they are looking for on a web-site within two clicks, otherwise they will become frustrated and leave. The applications developed for this project have further simplified this guideline by ensuring that single navigation levels are used.

A2.3 Vary the feedback

Speech provides a natural output mechanism for user interfaces, particularly in those designed for young children. However, applications become tiresome if the same feedback is continually used. To engage the user and provide a more natural interface, the application should provide varied feedback in the same way a teacher would. For example, for a correct response, the feedback could be selected from a set of words and phrases appropriate to the level of progress such as: "Very good"; "Wow"; "Excellent" etc. Speech output has been implemented in all the applications which concatenate random phrases, chosen from sets of pre-recorded utterances, into spoken sentences. Chiasson and Gutwin (2005) report an experiment that was carried out to establish whether children were significantly affected by praise given by a computer system; the researchers reported their surprise when the results suggested that there was no statistically significant finding to suggest that children were affected significantly by praise. This finding suggested that an evaluation to establish whether praise is necessary at all in a computer interface may be worthy of a follow up study.

A2.4 Minimise keyboard input

Depending on the type of activities to be carried out on a computer, the use of a standard "QWERTY" keyboard may cause problems for children. For example, all the letters are upper case, which is fine if the activity is to teach letter names (which are taught in upper case), but not if the activity is to teach grapheme symbols (which are taught using groups of lower case letters). The character rendered when a key displaying an upper case character is pressed is a lower case character, which is confusing whatever the activity. Some lower case characters produce different shaped characters on the screen to those used in the classroom and the lower case characters produced on the screen can vary depending on the font being used. The characters are not presented on the keyboard in alphabetical order which is very confusing to children trying to locate them. Accidentally pressing “caps-lock”, “shift” or “shift-lock” can change the display from lower case to upper case which causes further confusion.

Due to the problems associated with the use of “QWERTY” keyboards by young children, this guideline recommends that this choice of input mechanism is avoided and if possible not used at all. None of the applications developed for this study use the QWERTY keyboard as an input device. It is recommended that more sophisticated input devices such as a touch-screen and graphic tablet and pen are considered where funding for such hardware is available. However, developers of systems designed for classroom use need to be mindful that many primary classrooms will only have access to a standard QWERTY keyboard and mouse. To avoid keyboard use, alphabetic characters can be provided on the screen by use of graphics matching those used in the classroom. When designing the size of graphical letters for the screen, it is important to ensure that the size of the graphics is appropriate for the age of user. Hourcade et al (2007) reported that the 4 and 5 year old children in their study, despite being frequent mouse users, had low accuracy rates when pointing and clicking on the smallest target that they tested (16 pixels in diameter). Similarly Druin et al (2004) found in a study comparing the effectiveness of a mouse when used by 4 year olds, 5 year olds and adults, that age and target size had a significant effect on accuracy. The minimum target size for graphics on the Letterworld screen is 50 pixels to ensure there are no usability issues associated with clicking the desired on-screen components.

A2.5 Make use of randomness

Hanna et al (1997) claimed that "challenge" was one of the dimensions of engagement that was linked to the likeability and usability of computer applications. Randomness could improve the application in the way that it maintains challenge such as asking the questions in a random order to avoid the child learning the sequence of questions and associated answers as opposed to the challenge of working out the answers each time.

Jacoby (1978) claims that computer applications should randomise events that may be learnt as a sequence where the sequence is not an objective. Observations during evaluation sessions revealed that young children quickly learn sequences. It is therefore important to ensure that children cannot successfully complete activities by remembering what they did last time by, for example, varying the order in which children are asked questions. However, if the sequence is the objective of the exercise, as for example in a letter formation activity, the sequence should be the same each time.

A2.6 Enable early exit from the application which is inaccessible to children

Designers of computer applications for children need to ensure that the interface does not allow early termination of the session by the child intentionally or unintentionally. A child who likes to "click around" computer screens may unintentionally terminate a session and a child who is finding an activity difficult may terminate the session intentionally if there is a facility available for them to do so; the outcome of either of these scenarios would be worsened if the application was not recording details from which the teacher could establish how much of the session had been completed by the child. However, a teacher wishing to terminate the activity early should be able to do so. After related discussion with teachers on this topic, the suggested method of enabling early exit is to input a key combination that is unlikely to be guessed by the child. This guideline is specifically for use in educational applications where the parent or teacher needs to know that a session has been completed by a child.

A2.7 Provide tiered pro-active interactive help

This is a non-trivial task. Various methods of providing help have been devised and evaluated in applications tested in the classroom. The results of this work are summarised here:

Help controls for early readers should provide assistance by some means other than text; this could take the form of speech output, sound effects, automatic cursor movement which interacts with controls or a combination of these features. Only one help control should be provided for each screen; young children can become confused by more than one manual help source. The help control should offer good affordance and have consistent positioning across screens so that the user can easily locate it. Help should be available manually using a help control or automatically generated by the computer as a result of monitoring the child's progress and error patterns if required.

Tiered help may be provided if required. This involves providing high level hints at Tier 1, moving down the tiers offering more detailed help at each stage until it eventually provides the answer. Each tier should be designed to encourage the child to think the problem through and try to solve it only to be told the answer as a last resort. If the answer is given, the question should be rescheduled to be asked again later in the activity. Help can be proactive. This requires the child's progress to be monitored and responding to accordingly. For example, evaluation observations have shown that some children do not like to attempt a question if they aren't confident that they can answer it. With pro-active help, the application would not necessarily wait for the child's help request but would intervene by prompting the child through tiered help if, for example, they had made no attempt to answer the question within a specified time. Help can be interactive. When the child needs an answer, process or method explaining, the application should, where possible, show them. For example, the child cannot decide whether to select the "b" or "d" character and after the prompting and advice has been exhausted, the child could be shown the process to arrive at the answer by automatically moving the cursor to the correct letter, click it, say the letter's name and sound then relinquish control back to the user.

Computer application designers should evaluate the various forms of help discussed in this section to make an informed choice of the type and level of help to implement.

A2.8 Enable user defined options but make them inaccessible to the children.

User defined options should be hidden from children. Flexibility in computer programmes for children can be provided in many ways. Some applications provide options to customise content and pace whereas others may provide options to change the application's content due to a change in curriculum for example.

There are various implementation methods for application customisation: the simple method provided by Letterworld is illustrated in Figure A-6. It enables teachers to: choose a subset of letters to be attempted; choose click or drag mode; choose whether the child should be asked to find objects in the picture after completing the letter formation; choose whether instructions should be provided; choose the number of times the activity is to be repeated. However, more sophisticated techniques enabling individual profile settings can be provided using a database with the most sophisticated adaptation method requiring the least attention from the teacher being an Intelligent Tutoring System (ITS); this uses an expert knowledge base to adapt automatically to the user's requirements.

Flexibility to curriculum requirements has been implemented in the phonics system developed for this thesis and is discussed in Appendix E - *SPEL Application Configuration*.

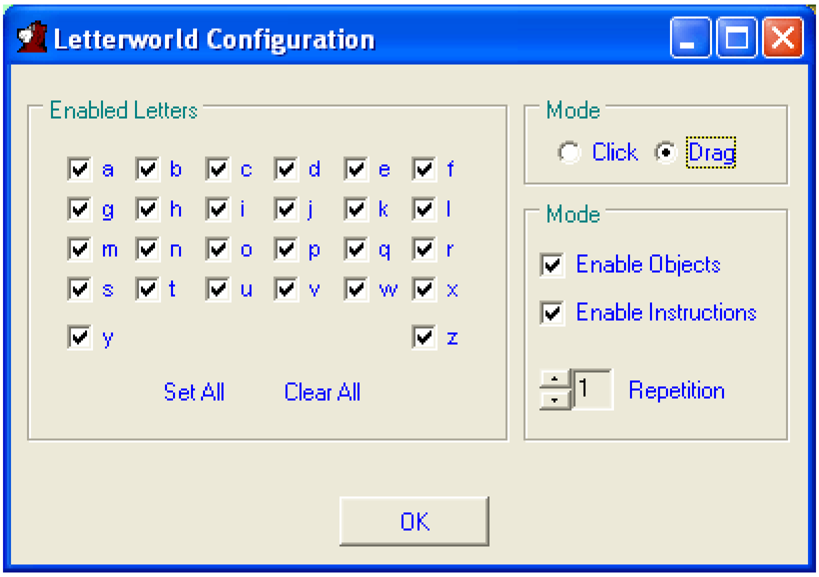

Figure A-6: User defined options interface used in the Letterworld prototype

Evaluation of the configuration facilities implemented in "Letterworld" found it to be too restrictive as it needs to be modified each time a different child uses the application which places an increased burden on the teaching staff. This method may however, be acceptable in a home setting where a single user is typical.

A2.9 Ensure that interaction hardware is of an appropriate size for the user

Young children are little people so interaction hardware should be sized accordingly. A small mouse or small digital pen for example, should be considered when specifying hardware requirements. Although Hourcade et al (2007) report from an evaluation that *"no statistically significant differences in accuracy or efficiency between the children who used small and regular-sized mice"*, the evaluation of an educational literacy application for Literacyworld, revealed one little girl whose hands were so small that she was unable to press the buttons without using two hands: one to hold the mouse and the other to press the button.

A2.10 Design for no scrolling

Young children often find it difficult to use a scroll control. It is recommended that computer programmes for this age group use a full screen, fixed size approach; this ensures that all areas of the activity are visible so cannot be hidden in an off-screen area as in the case of scrollable or resizable windows.

A3 Conclusions

The interface design guidelines developed for this project are not considered to be a definitive set but a useful working set for researchers and developers. The guidelines are particularly useful to developers with limited experience in working with young children and researchers wishing to improve, enhance add to or challenge the guidelines. The substantial number of usability evaluations conducted between Snape and Nicol in the development of the guidelines has enabled invaluable experience to be gained in working with young children from different schools and regions in a classroom setting. It is now clear that the evaluation of computer software with child-users in a classroom setting is an essential phase in software development; regardless of how well designed developers think the interface is, the children are likely to do something unexpectedly.

Appendix B

List of English Graphemes and Phonemes

Appendix B List of English graphemes and phonemes

The list of words for the SPEL activities is based on phases 2 to 5 of the DfES recommended phonics scheme “Letters and Sounds” (DfES, 2007b). The six word sets are used on alternate weeks to carry out segmenting and sequencing activities over a 12 week period.

Grapheme	Phoneme (IPA)	Sample words							
Phase 2 – Week 1									
<s>	/ s /	sat							
<a>	/ æ /	pat							
<t>	/ t /	tap							
<p>	/ p /	sap							
<i>	/ ɪ /	sit	pit	tip					
<n>	/ n /	nip	pin	tan					
<m>	/ m /	man	mat	map					
<d>	/ d /	did	dim	dip					
<g>	/ p /	tag	gap	pig					
<o>	/ ɒ /	got	top	pop					
<c>	/ k /	cap	cot	can					
<k>	/ k /	kid	kit	kip					
<ck>	/ k /	kick	sock	dock					
Phase 2 – Week 1									
<e>	/ e /	get	ten	peg					
<u>	/ ʌ /	run	mug	sun					
<r>	/ r /	rim	rug	rot					
<h>	/ h /	hot	him	hug					
<b>	/ b /	big	bet	bad					
<f>	/ f /	fit	fun	fog					
<ff>	/ f /	off	huff	cuff					
<l>	/ l /	lap	leg	lot					
<ll>	/ l /	fill	doll	sell					
<ss>	/ s /	less	boss	fuss					
Phase 3 – Week 2									
<j>	/ dʒ /	jug	jet	jog					
<v>	/ v /	van	vat	vet					
<w>	/ w /	win	web	wig					
<x>	/ ks /	tax	box	fix					
<y>	/ j /	yap	yes	yet					
<z>	/ z /	zip	zap	zit					
<zz>	/ z /	buzz	jazz	fizz					
<qu>	/ kw /	quiz	quit	quick					
Phase 3 – Week 2									
<ch>	/ tʃ /	chop	chin	much	chug	such	rich		
<sh>	/ ʃ /	shed	shop	cash	shed	hush	bash		
<th>	/ ð /	the	this	with	then	that	thus		
<ng>	/ ŋ /	ring	song	king	hang	rung	wing		
Phase 3 – Week 3									
<ai>	/ eɪ /	wait	tail	rain	main				
<ee>	/ iː /	feel	deep	week	weep				
<igh>	/ aɪ /	high	sigh	light	night				
<oa>	/ əʊ /	oak	coat	soap	toad				

<oo>	/ uː /	zoo	boot	food	hoof				
<ar>	/ ɑː /	car	bark	hard	park				
<or>	/ ɔː /	sort	born	fort	cord				
<ur>	/ ɜː /	burn	hurt	curl	turn				
<ow>	/ aʊ /	now	owl	how	down				
<oi>	/ ɔɪ /	oil	boil	coin	coil				
Phase 5 – Week 4									
<ear>	/ ɪə /	dear	fear	near	beard				
<air>	/ eə /	fair	lair	pair	cairn				
<ure>	/ ʊə /	pure	cure	lure	mature				
<er>	/ ə /	better	ladder	dinner	supper				
<ay>	/ eɪ /	may	play	tray	spray				
<ou>	/ aʊ /	cloud	found	proud	sound				
<ie>	/aɪ /	tie	cried	tried	fried				
<ea>	/ iː /	seat	read	beat	treat				
<oy>	/ ɔɪ /	boy	toy	joy	royal				
<ir>	/ ɜː /	sir	girl	bird	skirt				
Phase 5 – Week 5									
<ue>	/ uː /	blue	clue	glue	true				
<aw>	/ ɔː/	saw	raw	claw	lawn				
<wh>	/ w /	when	wheel	which	whack				
<ph>	/ f /	dolphin	phonics	alphabet	phantom				
<ew>	/ uː /	grew	crew	brew	drew				
<oe>	/ əʊ /	toe	foe	woe	goes				
<au>	/ ɔː/	haul	daub	taut	launch				
<ey>	/ I /	donkey	jockey	chimney	turkey				
<a-e>	/ eɪ /	made	take	game	race				
<e-e>	/ iː /	gene	theme	these	extreme				
Phase 5 – Week 6									
<i-e>	/ aɪ /	like	ripe	time	pine				
<o-e>	/ əʊ /	woke	pole	note	home				
<u-e>	/ uː/	rule	rude	flute	prune				

Digraphs on week 6 screen must include the following to accommodate the revision word set below: <igh> <oi> <oo> <ay> <ou> <ew> <a-e>

		line	fight	soil	moon				
		robe	bite	hole	vote				
		cape	mole	screw	came				
		yule	dude	hive	nine				
		life	toil	cool	right				
		gate	rope	lobe	delay				
		bone	tape	bake	sprout				

Appendix C

Usability Evaluation Questionnaire

Appendix C Usability Evaluation Questionnaire

Post evaluation questions

Children point to Smileometer when required (Figure 4-1)

1 How good was the computer in teaching you phonics?
(If they don't seem to understand the term phonics, explain it)

2 How much would you like to use the programme in school?

3 Do you have a computer at home that you are allowed to use?

4 How easy was it to use the sequencing activity?
(Point to the screen shot to remind them which that was)

5 How easy was it to use the segmenting activity?
(Point to the screen shot to remind them which that was)

6 If you had a computer at home and were allowed to use it, would you want this programme on it?

Appendix D

Test score sheet and instructions

Appendix D Test score sheet and instructions

Child ID: School: Class:

Letter Sound Knowledge			
Grapheme		LN	Mark
1	s		
2	m		
3	p		
4	t		
5	i		
6	n		
7	a		
8	f		
9	h		
10	d		
11	j		
12	ee		
13	sh		
14	ch		
15	th		
16	oo		
17	ay		
	(17) Total:		

Sound Isolation		
Initial Phoneme		
Teaching Items:		
king saf rub san guf		
Say initial sound		Mark
1	bem	
2	mig	
3	feep	
4	swib	
5	brug	
6	drick	
	(6) Total:	
Final Phoneme		
Teaching Items:		
duck fass pag		
Say End Sound		Mark
7	baff	
8	kass	
9	wib	
10	zind	
11	besk	
12	doost	
	(6) Total:	
(12)Total Combined:		

Early Word Recognition			
Word		Mark	Mark
1	cat		
2	in		
3	up		
4	and		
5	off		
6	went		
7	duck		
8	but		
9	you		
10	was		
11	frog		
12	king		
13	help		
14	fish		
15	said		
16	then		
17	door		
18	school		
19	out		
20	orange		
21	what		
22	bread		
23	cake		
24	bird		
25	shoe		
26	dragon		
27	water		
28	fire		
29	biscuit		
30	giant		
(15) Exception Total:			
(15) Regular Total:			

Sound Deletion				
Spkr	Word	Delete	Expect	Mark
T	seesaw	saw	see	
T	ice-cream	ice	cream	
	starfish	fish	star	
	toothbrusl	tooth	brush	
T	sheep	p	she	
T	face	s	fay	
	boat	t	boe	
	house	s	how	
	desk	k	dess	
T	bed	b	ed	
T	goat	g	oat	
	shop	sh	op	
	parrot	p	arrot	
	cloud	e	loud	
	stamp	s	tamp	
T	frog	r	fog	
	mask	s	mak	
	drink	r	dink	
	jumper	m	jupper	
			(12) Total:	

Split Vowels	
Teaching Items:	
mat make bit bite	
Grapheme	Mark
1 o-e	
2 e-e	
3 a-e	
4 u-e	
5 i-e	
(5) Total:	

Digit Span Test							
Forward Digit Span				Reverse Digit Span			
Digits (a)	Mark	Digits (b)	Mark	Digits (c)	Mark	Digits (d)	Mark
43		16		83		29	
792		847		475		615	
5941		7253		2619		3852	
93872		75396		28736		59413	
152649		216748		624719		276391	
3745261		4925316		4183627		1586937	
(6) Total:		(6) Total:		(6) Total:		(6) Total:	
				(24) Total Combined:			

Pupil Score Sheet Pre Test

Figure D-1: **Pupil Score Sheet**

Appendix E

SPEL Application Configuration

Appendix E SPEL Application Configuration

The SPEL application has been designed to be configurable with regard to activity, word and grapheme lists.

The word lists and activities can change automatically by date if required to minimise supervision. The "trigger" dates to change activity can be changed by modifying the contents of the text file "WordListChangeDates.txt".

An example of the file with test dates is shown in Figure E-1. When the experiment was carried out, this file held the trigger dates for each of the 12 weeks to ensure all activities were changed at the same time in all schools on a weekly basis. If a problem prevented a class from working on the computer, the teaching staff or school technician would have contacted the researcher to determine the best way forward. The most practical approach would have been to modify the trigger dates to accommodate any missed classes if necessary. However, if a child had missed a whole week, they could catch up by overriding the date as illustrated in Figure E-8 on page E-7. However, the application was trouble free across all schools for the duration of the trial.

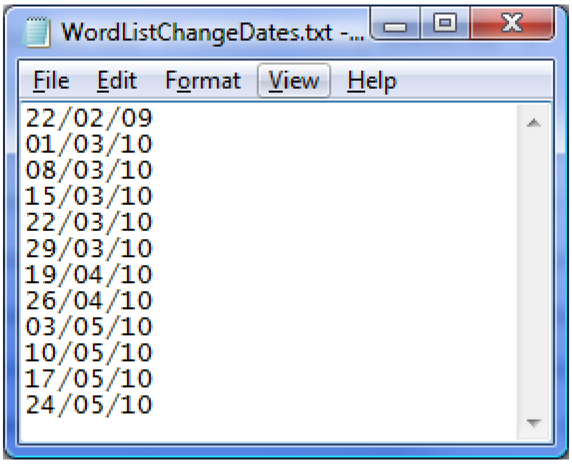

Figure E-1: Activity / word trigger dates

The dates trigger the activity to load the Sequence or Segment activity and the words and grapheme lists.

The activity sequence can be easily changed by modifying the file "Activity.txt". The file used in the experiment is illustrated in Figure E-2 showing the activities alternating on each trigger date.

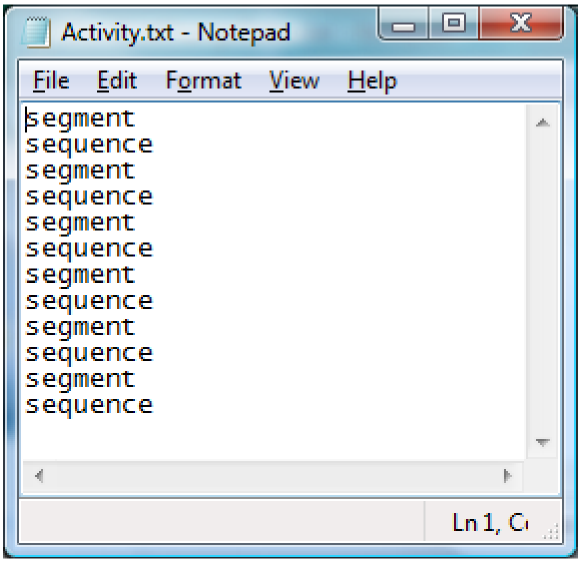

Figure E-2: Activity sequence based on date triggers

The word lists loaded in sequence are determined by their filenames. Words0.txt is the first file to be loaded then after a trigger date Words1.txt will be loaded and so on. So for the experiment there were 6 files of words which were used for segmenting and sequencing over the 12 week experimental period.

Figure E-3 and Figure E-4 provide examples of parts of two word sets used in the experiment.

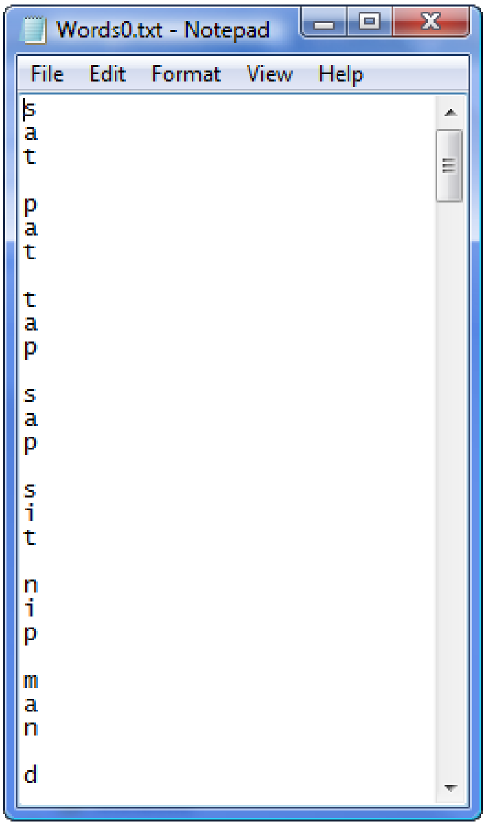

Figure E-3: Words0 list

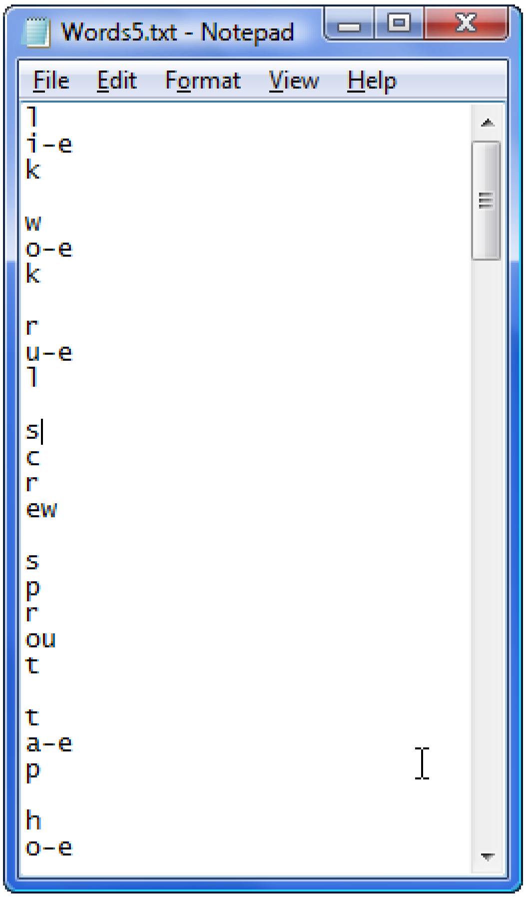

Figure E-4: Words5 list

To modify, add or remove words from the list is simply a case of editing the appropriate text file. It should be noted that each word is presented in its grapheme format (one grapheme per line) with a blank line to separate words.

If a new word list needs to be added, a new word file can be created following the sequential file naming convention and words added in the format shown. Note: each new word will need the corresponding wave files for the word and graphemes. Many more word sounds have been recorded than are required by the experiment for flexibility but if a word is required that is not present in the recorded vocabulary, the word can be recorded and added to the default installation directory: "c:\program files\UCLAN\SPEL\sounds\word". The filename of the recorded word must be the same name as the word with a ".wav" file extension. For example, the word "Hello" would be recorded and saved in the \sounds\word directory as "Hello.wav". This simple format is illustrated in Figure E-5.

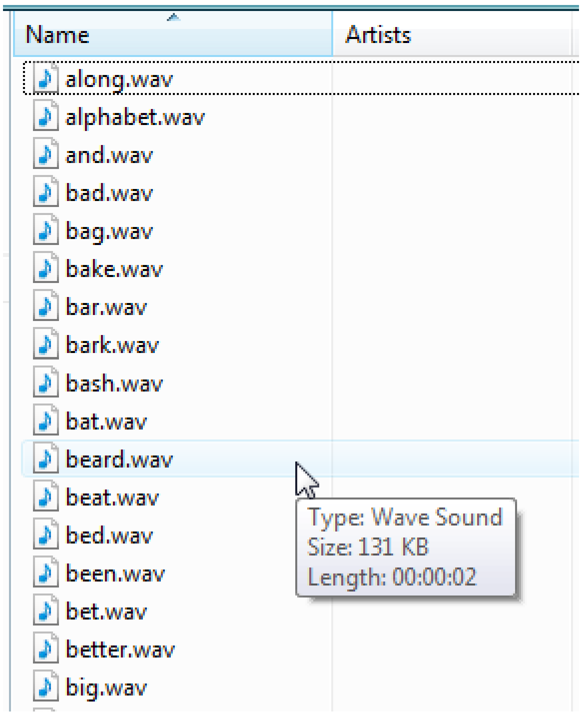

Figure E-5: Partial listing of the \Word directory

The grapheme lists to be loaded are also in text files that follow the same naming convention. For example, for words0.txt there would be an accompanying grapheme file called “words0_digraph.txt”. Note: the alphabet is present on all screens so only extra graphemes such as digraphs need to be present in these files. Figure E-6 and Figure E-7 show the digraph lists for word sets 0 and 5.

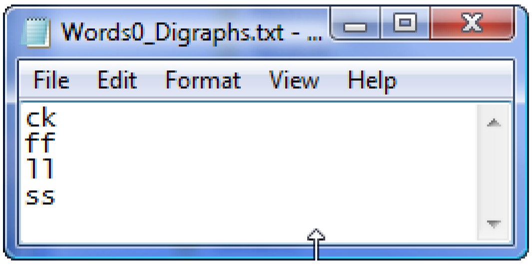

Figure E-6: Words0 digraphs

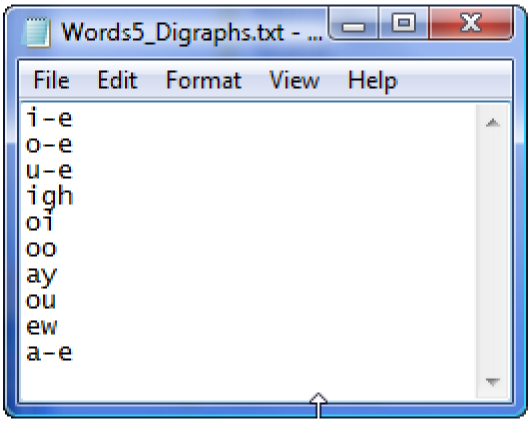

Figure E-7: Words5 digraphs

If other word or grapheme files need to be provided, new files can be created using the same filename convention and store the files in directory \wordlists.

If the teacher wishes to override the trigger date or wishes to quickly check which words are to be displayed in a particular set, holding down the shift and Ctrl keys whilst right-clicking on Floppy Dog will bring up the dialog box in Figure E-8.

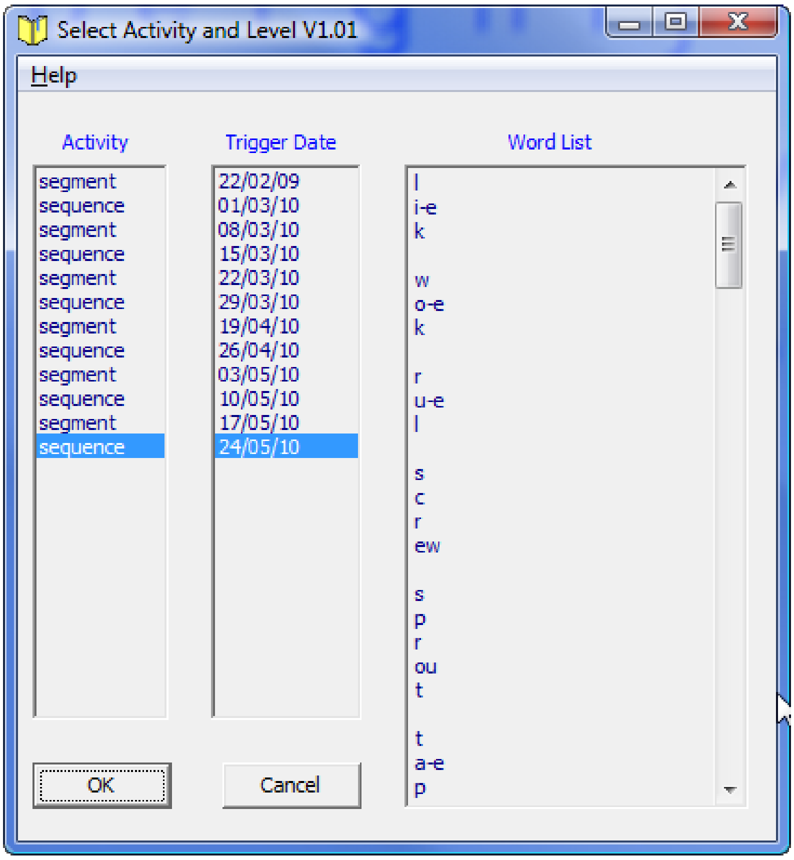

Figure E-8: Teacher activity override screen

The key combination has been chosen to avoid a child inadvertently entering the administration screen. Clicking on any of the activities in this dialog box will display the associated word list. If used as a reminder, cancel should be clicked when finished. However, if a specific activity and word list is required regardless of the trigger date, select the required activity and word list as shown and click the OK button; the selected word list and activity will then start.

If the application is considered useful to teachers, the wordlist, dates, phoneme lists and voice recording could be incorporated into this screen to avoid having to modify individual text files. However, this is not required for the experiment so could be an area of further work.

The technicians at each school were given the instructions to download SPEL from the Internet. They only needed to run the single .msi file to install all the files and it will automatically create the appropriate directory structure illustrated in Figure E-9. The user can change the target directory during installation but the default target is drive C and in program files the directory structure illustrated in Figure E-9 is automatically created.

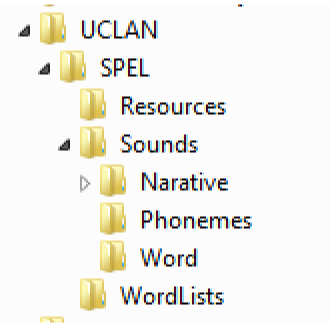

Figure E-9: SPEL directory structure

The technicians replicated the application across the school network onto PCs and Laptops to be used by the intervention groups. Once the application was installed, the researcher visited the schools to check the integrity of the installation before the experiment started.

Appendix F

Random Boolean Algorithm

Appendix F Random Boolean Algorithm

Using the algorithm: `Int((upperbound - lowerbound + 1) * Rnd + lowerbound)`

In this case, `upperbound is 1 and lowerbound is 0 so 1-0+1=2`

This simplifies the function to : `Int(2*Rnd)`

The following VBA macro creates the set of random Boolean numbers

```
Sub Rand_Bool()
' based on Int((upperbound - lowerbound + 1) * Rnd + lowerbound)

 Randomize  ' Seed RNG with random value from system clock

 Dim Row As Integer

 'Create set of random Booleans and drop into convenient place on sheet
 For Row = 40 To 75 ‘Drop into spreadsheet at convenient location
   ActiveSheet.Cells(Row, 2) = Int((2) * Rnd)
 Next Row
End Sub
```

Appendix G

Capturing interaction and performance data

Appendix G Capturing interaction and performance data

G1 Overview

For research purposes, the SPEL application was designed and implemented to collect results into paradox database tables; an example is illustrated in Figure G-1.

.

Table : D:\Work (Z)\...\Capture.DB

Capture	IdNumber	SessionNum	Object	DownTime	DownX	DownY	DownString	UpTime	UpX	UpY	UpString	Button
1	2	1	lbC	14:28:28	15	51	c	14:28:29	15	52	c	Left
2	2	1	lbA	14:28:45	11	50	a	14:28:46	11	50	a	Left
3	2	1	lbT	14:28:58	10	44	t	14:28:58	10	44	t	Left
4	2	1	lnV	14:29:12	4	44	v	14:29:12	4	46	v	Left
5	2	1	lbA	14:29:14	10	46	a	14:29:14	10	78	a	Left
6	2	1	lbN	14:29:16	14	40	n	14:29:16	16	40	n	Left
7	2	1	lbB	14:29:27	6	40	b	14:29:27	7	40	b	Left
8	2	1	lbB	14:29:29	7	40	b	14:29:29	7	40	b	Left
9	2	1	lbI	14:29:32	8	36	i	14:29:33	-16	50	i	Left
10	2	1	lbG	14:29:34	14	50	g	14:29:34	14	50	g	Left
11	2	1	lbH	14:29:47	15	60	h	14:29:47	11	40	h	Left
12	2	1	lbO	14:29:49	11	55	o	14:29:49	11	55	o	Left
13	2	1	lbT	14:29:51	7	43	t	14:29:51	8	38	t	Left
14	2	1	lbH	14:30:04	23	59	h	14:30:04	22	59	h	Left
15	2	1	lbA	14:30:05	16	47	a	14:30:05	58	71	a	Left
16	2	1	lbT	14:30:07	14	53	t	14:30:07	15	53	t	Left
17	2	1	lbL	14:30:19	6	33	l	14:30:19	6	33	l	Left
18	2	1	lbI	14:30:26	9	37	i	14:30:26	8	38	i	Left
19	2	1	lbP	14:30:27	18	46	p	14:30:27	18	46	p	Left
20	2	1	lbB	14:30:34	3	53	b	14:30:34	5	53	b	Left
21	2	1	lbA	14:30:35	22	50	a	14:30:35	22	50	a	Left
22	2	1	lbG	14:30:37	7	64	g	14:30:37	7	64	g	Left
23	2	1	lbS	14:30:47	16	48	s	14:30:47	16	48	s	Left
24	2	1	Image1	14:30:48	496	72	Background	14:30:49	496	78	Background	Left
25	2	1	lbI	14:30:53	5	42	i	14:30:53	-173	194	i	Left
26	2	1	lbT	14:30:55	8	33	t	14:30:55	8	33	t	Left
27	2	1	lbS	14:31:04	16	54	s	14:31:04	26	54	s	Left
28	2	1	lbI	14:31:07	3	44	i	14:31:07	5	45	i	Left
29	2	1	lbX	14:31:09	10	38	x	14:31:09	7	40	x	Left
30	2	1	lbH	14:31:18	2	42	h	14:31:19	0	42	h	Left
31	2	1	lbO	14:31:20	10	33	o	14:31:20	35	34	o	Left
32	2	1	lbT	14:31:21	5	38	t	14:31:21	5	38	t	Left

Record 30 of 96

Figure G-1: Example data capture table

Using these data recorded into the tables, the child's interaction session can be replayed at any speed and paused at will as discussed in Section 5.4.1. The table also stores other useful data which can be used in further analysis if required. Mouse move coordinates are held in a different table as there are many more of them – the granularity of the illustrated table is at mouse-click level.

An overview of the table fields is provided here:

G1.1 Capture

This is the mouse event capture number. This column is useful to identify how many times the child has clicked the mouse and is useful for discussion purposes to identify a particular event.

G1.2 Id Number

This is the ID number given to the child to maintain anonymity.

G1.3 Session Number

This identifies the session; for example, sequencing session 1, segmenting session 3 etc., which can be used to create a unique identifier (Capture/IdNumber/SessionNum) for mouse clicks when referring to multiple sessions.

G1.4 Object

This is the object reference on the screen that the child has clicked. For example, lbC is the object associated with grapheme <c>. This is useful as it can be compared to the grapheme that the child should have clicked which is held in a different table.

G1.5 Downtime

This is the time recorded when the child pressed the mouse button down. This is useful when used with the UpTimevalue to determine how long the child kept the mouse button pressed. In conjunction with the X,Y coordinates this can indicate problems such as drag and drop usability issues.

G1.6 DownX and DownY

These are the X and Y co-ordinates of the screen position when the mouse button is pressed. This information is useful in determining that the child clicked an incorrect grapheme or other control. For example, capture number 24 in Figure G-1 shows the clicked object to be the background. It is possible that the child could have just missed the letter or perhaps they were messing about and clicked the screen elsewhere; by looking at the mouse co-ordinates, it is obvious that they are not near any letters so must have been messing about and this result can probably be discounted as being a near-miss on a grapheme. This type of result can be used as an interface error indication; if the child had just missed the grapheme, the possibility that the graphics are not large enough or the target area is too precise would have to be considered as an interface issue if it happened regularly, particularly with more than one child.

G1.7 DownString

This is the text string of the grapheme the child actually clicked. This is used in textual comparisons for correctness and to identify an associated sound file.

G1.8 UpTime

This is the time recorded when the child released the mouse button. This is useful when used with the DownTime recording for the same capture, as it shows whether the child hesitated over a choice.

G1.9 UpX and UpY

These are the X,Y co-ordinates on the screen of when the child released the mouse button. This data is useful to determine whether the child has dragged the mouse cursor and why. For example, if the cursor has been dragged a long way, it could suggest that the child thinks that the letter should be dragged and this may have to be considered as an interface issue. However, if the mouse has been dragged a short distance, it is more likely that the hand of the child has just slipped slightly; a difference in the down and up coordinates is typical of children struggling to manage the mouse – particularly if the mouse is too large for them.

G1.10 UpString

This is the text string of the object over which the mouse button was released. This is useful when used with the DownString data because if the down string and up string are different, it means the child may have attempted to drag one letter over another and this may need to be considered as an interface issue.

G1.11 Button

This shows which mouse button was pressed by the child. It is useful to ensure the children are not trying to use the right mouse button. If this had been an issue, a solution such as programming both buttons as the action button may have been considered; in all the evaluation tests this was not an issue even with left-handed children.

Figure G-2 illustrates another set of results and is included here as a comparison to Figure G-1. This is data collected from one of the stronger members of the group. This data shows that the child has not clicked on the background at all and has made use of Floppy Dog as a prompt. It is also a good example of a child gaining confidence with the application as there are substantial delays between the choice of the first few graphemes then the child speeds up and uses help less as time progresses. This child continued to grow in confidence and eventually managed to get most of the long vowel digraph sessions correct.

Table : D:\Work (Z)\...\Capture.DB

Capture	IdNumber	SessionNum	Object	DownTime	DownX	DownY	DownString	UpTime	UpX	UpY	UpString	Button
1	2	1	lbFloppy	11:05:49	58	72		11:05:51	58	72		Left
2	2	1	lbC	11:06:05	21	50	c	11:06:05	21	50	c	Left
3	2	1	lbA	11:06:08	14	52	a	11:06:08	14	52	a	Left
4	2	1	lbT	11:06:10	12	54	t	11:06:10	12	54	t	Left
5	2	1	lbFloppy	11:06:20	63	74		11:06:21	63	74		Left
6	2	1	lnV	11:06:25	18	57	v	11:06:25	18	57	v	Left
7	2	1	lbA	11:06:28	26	46	a	11:06:28	26	46	a	Left
8	2	1	lbN	11:06:33	15	51	n	11:06:33	15	51	n	Left
9	2	1	lbB	11:06:40	20	56	b	11:06:40	20	56	b	Left
10	2	1	lbI	11:06:45	13	[illegible]	i	11:06:45	13	61	i	Left
11	2	1	lbG	11:06:48	19	71	g	11:06:48	19	71	g	Left
12	2	1	lbH	11:06:58	19	62	h	11:06:59	19	61	h	Left
13	2	1	lbU	11:07:01	34	61	u	11:07:02	34	61	u	Left
14	2	1	lbT	11:07:03	7	52	t	11:07:03	7	52	t	Left
15	2	1	lbO	11:07:11	16	56	o	11:07:12	16	56	o	Left
16	2	1	lbT	11:07:14	11	58	t	11:07:14	11	58	t	Left
17	2	1	lbH	11:07:25	17	59	h	11:07:26	20	53	h	Left
18	2	1	lbA	11:07:30	7	50	a	11:07:30	7	50	a	Left
19	2	1	lbT	11:07:33	18	69	t	11:07:34	18	69	t	Left
20	2	1	lbL	11:07:41	12	59	l	11:07:41	12	59	l	Left
21	2	1	lbI	11:07:43	2	67	i	11:07:44	2	67	i	Left
22	2	1	lbP	11:07:50	20	58	p	11:07:50	20	58	p	Left
23	2	1	lbB	11:07:58	28	49	b	11:07:58	28	49	b	Left
24	2	1	lbA	11:08:00	20	44	a	11:08:00	20	44	a	Left
25	2	1	lbG	11:08:03	25	77	g	11:08:03	25	77	g	Left
26	2	1	lbS	11:08:11	10	46	s	11:08:11	10	46	s	Left
27	2	1	lbI	11:08:13	5	52	i	11:08:13	5	52	i	Left
28	2	1	lbT	11:08:15	14	39	t	11:08:15	14	39	t	Left
29	2	1	lbS	11:08:23	28	53	s	11:08:23	28	53	s	Left
30	2	1	lbI	11:08:26	3	59	i	11:08:26	3	59	i	Left
31	2	1	lbX	11:08:28	3	42	x	11:08:29	9	63	x	Left
32	2	1	lbH	11:08:40	26	59	h	11:08:40	26	59	h	Left

Record 1 of 113

Figure G-2: Data capture from a stronger pupil

Appendix H

Statistical Analysis Details

Appendix H Statistical Analysis Details

H1 Overview

This appendix provides detail to support the statistical analysis results reported in Section 6.3.2. It provides more of the SPSS generated data and the program code required to regenerate the results should the tests need to be replicated.

H2 Check for normality of the sample

A check for the suitability of subject distribution is considered here. The groups are allocated randomly grouped on age and gender as discussed in section 6.3.2.3. A histogram (Figure H-1) and Q-Q plot (Figure H-2) were created to observe the distribution of the test subjects. The histogram illustrates the normality of the whole sample which is seen to be reasonable. The Q-Q plot also shows a reasonable level of normality within the group. A normal distribution is generally considered a pre-requisite of variance analysis; however (Schmider et al., 2010) show that the ANOVA is actually tolerant of deviations from normality.

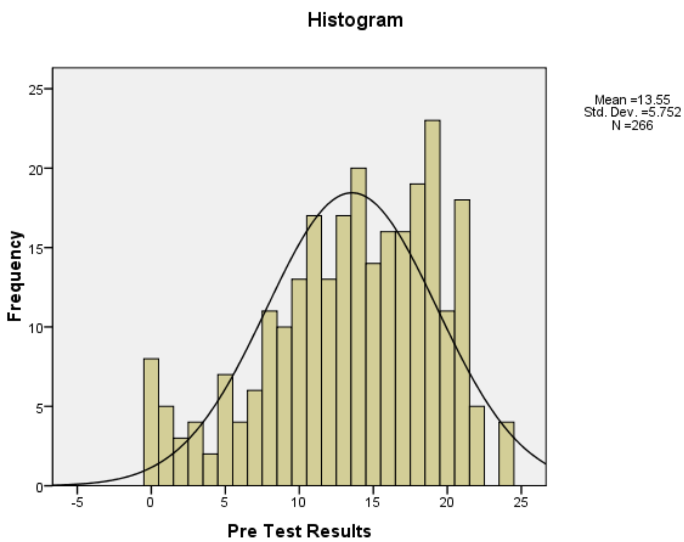

Figure H-1: Histogram of all participants' pre-test scores

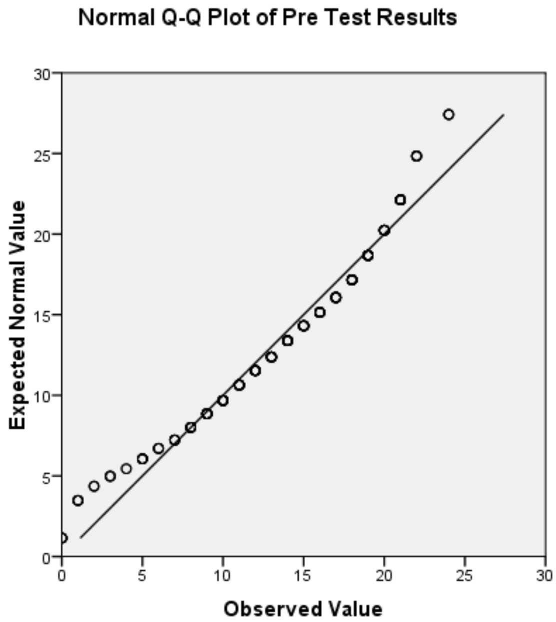

Figure H-2: Q-Q plot for control group pre-test scores

H3 Test Means and Standard Deviations

Report

GroupedPost		Pre Test Results	Post Test Results
Control	Mean	13.62	16.51
	N	133	133
	Std. Deviation	5.953	6.036
Intervention	Mean	13.49	16.26
	N	133	133
	Std. Deviation	5.566	5.315
Total	Mean	13.55	16.38
	N	266	266
	Std. Deviation	5.752	5.678

Table H-1: Pre-and Post-test means and standard deviations

Table H-1 shows the means and standard deviations of control and intervention groups to be well matched on ability before the trial started and shows minimal difference between the groups after the trial.

H4 ANCOVA

Table H-2 details the results of an ANCOVA on post-test control and intervention group differences in means using pre-test results as a covariate. The results show no statistical significance in between group scores ($p > 0.7$). The covariate ResultsPre is shown to be significant ($p < 0.001$) so justifies its inclusion as a covariate.

Tests of Between-Subjects Effects

Dependent Variable:Post Test Results

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[b]
Corrected Model	4768.223[a]	2	2384.112	166.113	.000	332.226	1.000
Intercept	1653.942	1	1653.942	115.239	.000	115.239	1.000
ResultsPre	4763.877	1	4763.877	331.924	.000	331.924	1.000
GroupedPost	1.732	1	1.732	.121	.729	.121	.064
Error	3774.664	263	14.352				
Total	79942.000	266					
Corrected Total	8542.887	265					

a. R Squared = .558 (Adjusted R Squared = .555)

b. Computed using alpha = .05

Table H-2: ANCOVA on post-test scores using pre-test scores as a covariate

H5 Independent samples T-test

As both groups are evenly matched at pre- and post-test, both groups progressed at the same rate. To test the statistical significance of this improvement, a T-test was carried out on the pre- and post-test score means. Table H-3: shows an increase in pre-test to post-test mean score of approximately 3 marks and Table H-4 shows this difference to be significant at ($p < 0.001$).

Group Statistics

	PrenPost	N	Mean	Std. Deviation	Std. Error Mean
AllResults	Posttest	266	16.38	5.678	.348
	Pretest	266	13.55	5.752	.353

Table H-3: Pre- and post-test means and standard deviations

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
AllResults	Equal variances assumed	1.513	.219	5.712	530	.000	2.831	.496	1.857	3.804
	Equal variances not assumed			5.712	529.911	.000	2.831	.496	1.857	3.804

Table H-4: T-test comparing Pre- and Post-test scores

H6 Effect size calculations

The calculation used to determine Cohen's *d* explained by (Becker, 2000) is:

$$d = \frac{\overline{x}_1 - \overline{x}_2}{s}$$

Where *d* is the difference of the post- and pre-test means divided by the standard deviation. As the populations of control and intervention groups are normal, it should not matter which standard deviation is used as they should both be the same. However, there are likely to be differences in reality so the pooled standard deviations can be used which uses standard deviations from both groups. The pooled standard deviation is determined by calculating the Root Mean Square (RMS) of the two standard deviations. The calculation used to determine the effect size in this trial (Cohen's *d*) is therefore:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}}$$

The between-groups effect size, calculated as the difference in gains divided by the pooled post-test standard deviation, was -0.02, which confirmed the non-significant difference between the groups.

H7 Attrition test

To determine whether the attrition significantly affected the results, an extreme case was analysed; values that were set to zero because a test had not been completed due to the child leaving the school were overridden and set to the maximum score of 24 in all 7 cases. The results show a small change in post-test means as expected but the analysis of covariance still shows the difference in the post-test means to be non-statistically significant ($p > 0.69$) so attrition has not affected the outcome of the trial.

Report

GroupedPost		Pre Test Results	Post Test Results
Control	Mean	13.62	17.23
	N	133	133
	Std. Deviation	5.953	5.417
Intervention	Mean	13.49	16.98
	N	133	133
	Std. Deviation	5.566	4.641
Total	Mean	13.55	17.11
	N	266	266
	Std. Deviation	5.752	5.036

Figure H-3: Means after extreme adjustment for attrition

Tests of Between-Subjects Effects

Dependent Variable:Post Test Results

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Noncent. Parameter	Observed Power[b]
Corrected Model	2980.990[a]	2	1490.495	104.811	.000	209.622	1.000
Intercept	3431.283	1	3431.283	241.287	.000	241.287	1.000
ResultsPre	2976.644	1	2976.644	209.317	.000	209.317	1.000
GroupedPost	2.182	1	2.182	.153	.696	.153	.068
Error	3740.063	263	14.221				
Total	84550.000	266					
Corrected Total	6721.053	265					

a. R Squared = .444 (Adjusted R Squared = .439)

b. Computed using alpha = .05

Figure H-4: ANCOVA results after extreme adjustment for attrition

H8 SPSS program code to replicate tests

The RCT results can be provided on request from the School of Computing, Engineering and Physical Sciences at the University of Central Lancashire.

To replicate the graphs and tables in this appendix, execute the following code within SPSS:

```
/* Create histogram with normal curve using all pre test scores */
FREQUENCIES VARIABLES=ResultsPre
  /STATISTICS=STDDEV
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.

/* Create group means */
MEANS TABLES=ResultsPre ResultsPost BY GroupedPost
  /CELLS MEAN COUNT STDDEV.

/* Create Q-Q plot of pre test scores */
PPLOT
  /VARIABLES=ResultsPre
  /NOLOG
  /NOSTANDARDIZE
  /TYPE=Q-Q
  /FRACTION=BLOM
  /TIES=MEAN
  /DIST=NORMAL.

/* Do ANCOVA using Post test results as dependant, contgrol and
intervention as independant and pretest scores as covariate */
UNIANOVA ResultsPost BY GroupedPost WITH ResultsPre
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /EMMEANS=TABLES(GroupedPost) WITH(ResultsPre=MEAN) COMPARE ADJ(LSD)
  /PRINT=DESCRIPTIVE OPOWER
  /CRITERIA=ALPHA(.05)
  /DESIGN=ResultsPre GroupedPost.

/* Independans samples t-test                                        */
/* Compares mean of all control and intervention post-test results */
/* to the mean of all control and intervention pre-test results    */
/* as ANCOVA has shown there to be no difference between groups    */
/* Data is stored in column "All results" with pre-test results    */
/* followed by post results for all schools. The independent       */
/* variable is PrenPost which are values 1 to mark pre-test and 2  */
/* to mark post-test vals                                          */
T-TEST GROUPS=PrenPost(2 1)
  /MISSING=ANALYSIS
  /VARIABLES=AllResults
  /CRITERIA=CI(.9500).
```

Appendix I

Results Spreadsheet Examples

Appendix I Results Spreadsheet Examples

The following spreadsheets are taken from the results Excel workbook. The three spreadsheets illustrate the data collected and reported on a single class in a single school. The first spreadsheet holds the pre-test data, the second holds post-test data and the third holds the YARC results calculated from the raw test results and YARC data tables. The workbook contains 50 spreadsheets: three for each of the 9 classes in the trial and the rest are lookup tables and final data prepared for import into the SPSS statistical analysis package.

The fields on the test score sheets are:

Child ID: unique ID provided by the school

Date of Birth: Provided by the school

Gndr: Gender of the child provided by the school

Months: the child's age in months on the day of the test:

```
=IF((DAY(PreTestDate)-DAY(C4)) < 0,TRUNC((YEAR(PreTestDate)-
YEAR(C4))*12+MONTH(PreTestDate)-MONTH(C4)-
1),TRUNC((YEAR(PreTestDate)-YEAR(C4))*12+MONTH(PreTestDate)-
MONTH(C4)))
```

Group: c or i for control or intervention allocated randomly

Pairs: Children are paired on age and gender so child 'a' is paired with another child 'a'. 'b' is paired with 'b etc. If there is an odd number of children the "Pairs" cell will be blank as that child is not paired but simply randomly allocated to a group

LSK: YARC letter Sound Knowledge test

SI(i): YARC Sound Isolation test – initial sound

SI(f): YARC Sound Isolation test – final sound

SI(tot): SI(i) + SI(f) total raw score for the sound isolation test

EWR(e): YARC Early Word recognition – phenomic exceptions

EWR(r): YARC Early Word recognition – phenomic regular

EWR(tot): EWR(e) + EWR(r) total raw score for Early Word Recognition

SD: YARC Sound Deletion test

SV: Split Vowel digraph test

DS(a to d): Digit Span forward and reverse tests

The remaining columns are the standardised scores computed from the YARC data.

The summary spreadsheet is more readable self-explanatory side by side comparison of pre- and post-test standardised scores to present to the school.

L1 pre-test

							Raw pre-Test Scores													
Child ID	Date of Birth	Gndr	Months	Group	Pairs		LSK	SI(i)	SI(f)	SI(tot)	EWR(e)	EWR(r)	EWR(tot)	SD	SV	DS(a)	DS(b)	DS(c)	DS(d)	DS(tot)
Q888283008001	21/10/2003	F	75	c	a		17	6	6	12	15	15	30	7	5	3	3	1	2	9
U888283008032	03/10/2003	F	76	c	b		17	6	6	12	15	15	30	8	2	3	3	1	1	8
E888283008029	08/12/2003	F	73	c	c		15	6	5	11	7	13	20	8	3	4	3	2	1	10
X888283008005	26/02/2004	F	71	c	d		16	3	3	6	7	11	18	2	2	3	3	0	0	6
P888283008008	06/05/2004	F	68	c	e		14	6	1	7	0	8	8	4	1	2	2	1	0	5
C888283008018	14/07/2004	F	66	c	f		15	5	5	10	15	15	30	7	5	3	3	2	2	10
F888283008022	22/07/2004	F	66	c	g		17	6	1	7	7	12	19	3	3	2	3	2	1	8
V888283008023	26/10/2003	M	75	c	h		16	6	5	11	7	11	18	5	5	4	3	1	1	9
L888283008035	19/11/2003	M	74	c	i		14	6	4	10	0	6	6	2	1	2	3	1	1	7
E888283008002	21/01/2004	M	72	c	j		12	6	4	10	5	13	18	5	3	3	2	1	1	7
Q888283008028	28/01/2004	M	72	c	k		17	5	5	10	14	14	28	9	5	3	2	1	2	8
E888283008031	02/02/2004	M	72	c	l		6	6	6	12	8	12	20	9	1	3	3	2	2	10
Z888283008016	04/03/2004	M	71	c	m		16	6	3	9	9	11	20	10	0	3	3	0	0	6
B888283008027	05/03/2004	M	71	c	n		17	6	5	11	7	14	21	9	1	2	1	2	1	6
L888283008006	07/04/2004	M	69	c	o		6	6	4	10	14	14	28	8	5	3	2	2	1	8
M888283008026	03/04/2004	M	70	c	p		15	5	4	9	12	15	27	9	2	3	3	3	3	12
P888283008010	16/05/2004	M	68	c	q		13	6	4	10	3	10	13	5	2	3	2	1	1	7
T888283008012	30/07/2004	M	66	c	r		11	4	0	4	2	9	11	1	3	2	3	0	0	5
J803231708113	18/08/2004	M	65	c			11	5	0	5	1	8	9	0	0	1	1	0	0	2
H888283008004	26/10/2003	F	75	i	a		16	6	4	10	7	11	18	7	1	3	3	1	1	8
R888283008019	10/10/2003	F	75	i	b		16	6	6	12	8	12	20	4	1	3	3	2	2	10
W888283008014	19/01/2004	F	72	i	c		17	6	6	12	14	15	29	9	1	4	4	2	1	11
F888341408013	26/02/2004	F	71	i	d		17	6	5	11	15	15	30	8	5	3	3	2	2	10
A888283008007	20/06/2004	F	67	i	e		14	6	5	11	7	11	18	5	3	2	3	2	2	9
H888283008033	18/06/2004	F	67	i	f		15	6	5	11	1	10	11	1	1	2	2	1	1	6
X888283008034	18/08/2004	F	65	i	g		17	6	1	7	6	13	19	3	1	3	2	3	2	10
Q888283008030	22/09/2003	M	76	i	h		17	6	6	12	15	15	30	12	5	4	4	2	2	12
K888358208003	14/11/2003	M	74	i	i		13	6	6	12	14	15	29	10	3	4	3	2	1	10
C888283008020	13/01/2004	M	72	i	j		6	3	2	5	10	14	24	6	5	3	3	1	1	8
R888283008021	03/01/2004	M	73	i	k		16	6	6	12	2	13	15	3	1	1	1	1	1	4
K888283008015	25/02/2004	M	71	i	l		11	5	5	10	0	3	3	2	0	3	3	2	1	9
D888283008009	26/03/2004	M	70	i	m		14	5	4	9	8	8	16	4	1	2	3	1	1	7
J888283008024	24/03/2004	M	70	i	n		17	6	6	12	15	15	30	2	5	2	2	1	0	5
G888283008013	27/04/2004	M	69	i	o		15	5	3	8	2	10	12	7	3	3	3	1	1	8
Y888283008025	06/04/2004	M	69	i	p		15	3	4	7	3	8	11	6	2	4	3	0	0	7
N888283008017	18/05/2004	M	68	i	q		16	6	6	12	14	14	28	5	2	2	2	1	1	6
U888283008003	28/08/2004	M	65	i	r		14	5	5	10	0	6	6	4	2	2	3	0	0	5

Test Date: 05/02/2010

YARK pre-Test Results									
Letter Sound Knowledge			Sound Isolation			Early Word Recognition			
Std Score	Ability	Age Equiv	Std Score	Ability	Age Equiv	Std Score	Ability	Age Equiv	Std Score
116	Above Average	6:10	118	Above Average	6:10	120	Above Average	7:01	99
116	Above Average	6:10	118	Above Average	6:10	120	Above Average	7:01	104
97	Average	6:06	106	Average	6:09	96	Average	6:05	104
110	Average	6:08	92	Average	5:07	108	Average	6:04	81
97	Average	6:04	97	Average	5:11	94	Average	5:10	95
103	Average	6:06	111	Average	6:07	>130	Outstanding	7:01	109
121	Excellent	6:10	97	Average	5:11	109	Average	6:05	89
105	Average	6:08	106	Average	6:09	94	Average	6:04	90
92	Average	6:04	98	Average	6:07	78	Severe Difficulty	5:08	72
83	Below Average	6:00	98	Average	6:07	94	Average	6:04	90
116	Above Average	6:10	98	Average	6:07	110	Average	6:10	108
<70	Very Severe Difficulty	4:09	118	Above Average	6:10	96	Average	6:05	108
110	Average	6:08	105	Average	6:04	110	Average	6:05	124
121	Excellent	6:10	119	Above Average	6:09	112	Average	6:06	119
<70	Very Severe Difficulty	4:09	111	Average	6:07	125	Excellent	6:10	114
103	Average	6:06	105	Average	6:04	122	Excellent	6:10	119
93	Average	6:02	111	Average	6:07	101	Average	6:02	100
85	Average	5:09	84	Below Average	5:03	99	Average	6:00	71
97	Average	5:09	95	Average	5:05	107	Average	5:11	<70
105	Average	6:08	98	Average	6:07	94	Average	6:04	99
105	Average	6:08	118	Above Average	6:10	96	Average	6:05	85
116	Above Average	6:10	118	Above Average	6:10	114	Average	6:11	108
121	Excellent	6:10	119	Above Average	6:09	>130	Outstanding	7:01	114
97	Average	6:04	119	Above Average	6:09	108	Average	6:04	100
103	Average	6:06	119	Above Average	6:09	99	Average	6:00	71
>130	Outstanding	6:10	103	Average	5:11	121	Excellent	6:05	95
116	Above Average	6:10	118	Above Average	6:10	120	Above Average	7:01	129
87	Average	6:02	118	Above Average	6:10	114	Average	6:11	113
<70	Very Severe Difficulty	4:09	75	Severe Difficulty	5:05	102	Average	6:08	95
105	Average	6:08	118	Above Average	6:10	90	Average	6:02	79
85	Average	5:09	111	Average	6:07	85	Average	5:05	81
97	Average	6:04	105	Average	6:04	105	Average	6:03	95
121	Excellent	6:10	130	Excellent	6:10	>130	Outstanding	7:01	81
103	Average	6:06	101	Average	6:01	100	Average	6:01	109
103	Average	6:06	97	Average	5:11	99	Average	6:00	105
110	Average	6:08	130	Excellent	6:10	125	Excellent	6:10	100
109	Average	6:04	118	Above Average	6:07	102	Average	5:08	101

Sound Deletion		Phoneme Awareness		
Ability	Age Equiv	Std Score	Ability	Age Equiv
Average	6:04	96	Average	6:09
Average	6:06	98	Average	6:10
Average	6:06	96	Average	6:09
Below Average	5:00	85	Average	5:04
Average	5:08	92	Average	5:10
Average	6:04	104	Average	6:07
Average	5:05	89	Average	5:08
Average	5:11	90	Average	6:06
Severe Difficulty	5:00	82	Below Average	6:00
Average	5:11	88	Average	6:04
Average	6:09	96	Average	6:09
Average	6:09	100	Average	6:11
Excellent	6:10	108	Average	6:09
Above Average	6:09	110	Average	6:10
Average	6:06	106	Average	6:08
Above Average	6:09	106	Average	6:08
Average	5:11	100	Average	6:04
Severe Difficulty	4:07	77	Severe Difficulty	4:11
Very Severe Difficulty	4:07	83	Below Average	4:11
Average	6:04	92	Average	6:07
Average	5:08	90	Average	6:06
Average	6:09	100	Average	6:11
Average	6:06	108	Average	6:09
Average	5:11	102	Average	6:06
Severe Difficulty	4:07	94	Average	6:00
Average	5:05	96	Average	5:08
Excellent	7:01	107	Average	7:01
Average	6:10	103	Average	6:11
Average	6:02	80	Below Average	5:10
Severe Difficulty	5:05	88	Average	6:04
Below Average	5:00	94	Average	6:00
Average	5:08	96	Average	6:02
Below Average	5:00	98	Average	6:03
Average	6:04	100	Average	6:04
Average	6:02	96	Average	6:02
Average	5:11	104	Average	6:07
Average	5:08	104	Average	6:03

L1 post-test

							Raw pre-Test Scores												
Child ID	Date of Birth	Gndr	Months	Group	Pairs		LSK	SI(i)	SI(f)	SI(tot)	EWR(e)	EWR(r)	EWR(tot)	SD	SV	DS(a)	DS(b)	DS(c)	DS(d)
Q888283008001	21/10/2003	F	79	c	a		17	6	6	12	15	15	30	12	5	3	2	3	3
U888283008032	03/10/2003	F	80	c	b		17	6	6	12	15	15	30	9	5	2	2	2	2
E888283008029	08/12/2003	F	78	c	c		6	6	4	10	15	15	30	10	5	5	5	2	2
X888283008005	26/02/2004	F	75	c	d		15	6	2	8	15	14	29	8	2	2	3	1	1
P888283008008	06/05/2004	F	73	c	e		15	6	2	8	4	11	15	5	1	2	1	0	0
C888283008018	14/07/2004	F	70	c	f		11	6	6	12	15	15	30	9	5	3	3	3	2
F888283008022	22/07/2004	F	70	c	g		17	6	3	9	11	15	26	7	3	2	3	1	1
V888283008023	26/10/2003	M	79	c	h		17	6	5	11	13	15	28	7	5	4	4	2	2
L888283008035	19/11/2003	M	78	c	i		13	6	6	12	0	6	6	2	1	3	3	1	1
E888283008002	21/01/2004	M	76	c	j		7	6	6	12	4	12	16	7	2	2	3	1	1
Q888283008028	28/01/2004	M	76	c	k		17	6	6	12	15	15	30	11	4	3	2	2	2
E888283008031	02/02/2004	M	76	c	l		16	6	6	12	5	13	18	10	4	2	2	1	2
Z888283008016	04/03/2004	M	75	c	m		17	6	3	9	12	14	26	10	3	4	3	2	3
B888283008027	05/03/2004	M	75	c	n		17	6	6	12	15	15	30	10	5	2	3	2	2
L888283008006	07/04/2004	M	74	c	o		17	6	5	11	14	15	29	10	4	4	3	1	2
M888283008026	03/04/2004	M	74	c	p		6	6	6	12	14	15	29	11	5	2	2	2	2
P888283008010	16/05/2004	M	72	c	q		16	6	6	12	10	15	25	9	5	2	2	1	1
T888283008012	30/07/2004	M	70	c	r		16	6	4	10	3	10	13	5	1	2	1	1	1
J803231708113	18/08/2004	M	69	c			15	2	1	3	6	8	14	3	4	2	2	0	0
H888283008004	26/10/2003	F	79	i	a		16	6	6	12	8	13	21	6	2	3	4	1	1
R888283008019	10/10/2003	F	80	i	b		17	6	6	12	13	15	28	8	5	5	4	2	1
W888283008014	19/01/2004	F	76	i	c		17	6	6	12	15	15	30	9	4	4	4	2	1
F888341408013	26/02/2004	F	75	i	d		17	6	6	12	15	15	30	12	5	4	4	2	1
A888283008007	20/06/2004	F	71	i	e		17	6	6	12	13	15	28	10	5	3	2	1	1
H888283008033	18/06/2004	F	71	i	f		14	6	6	12	1	11	12	4	0	3	3	1	2
X888283008034	18/08/2004	F	69	i	g		16	6	5	11	11	15	26	4	3	3	2	1	1
Q888283008030	22/09/2003	M	80	i	h		17	6	6	12	15	15	30	12	5	4	5	2	3
K888358208003	14/11/2003	M	78	i	i		6	6	6	12	15	15	30	12	5	4	4	4	3
C888283008020	13/01/2004	M	76	i	j		9	6	6	12	15	15	30	11	5	3	3	1	1
R888283008021	03/01/2004	M	77	i	k		15	6	6	12	7	13	20	8	5	4	2	3	2
K888283008015	25/02/2004	M	75	i	l		13	5	4	9	1	9	10	5	0	3	3	2	2
D888283008009	26/03/2004	M	74	i	m		17	6	0	6	12	13	25	6	2	3	3	2	2
J888283008024	24/03/2004	M	74	i	n		17	6	6	12	15	15	30	11	5	2	1	1	1
G888283008013	27/04/2004	M	73	i	o		17	6	6	12	7	15	22	3	3	2	2	1	1
Y888283008025	06/04/2004	M	74	i	p		15	6	4	10	3	11	14	8	3	5	4	2	1
N888283008017	18/05/2004	M	72	i	q		17	6	5	11	14	15	29	11	5	2	2	2	2
U888283008003	28/08/2004	M	69	i	r		14	6	6	12	2	10	12	8	1	3	3	1	1

Test Date: 11/06/2010

	YARK pre-Test Results								
	Letter Sound Knowledge			Sound Isolation			Early Word Recognition		
DS(tot)	Std Score	Ability	Age Equiv	Std Score	Ability	Age Equiv	Std Score	Ability	Age Equiv
11	112	Average	6:10	116	Above Average	6:10	115	Average	7:01
8	112	Average	6:10	116	Above Average	6:10	115	Average	7:01
14	<70	Very Severe Difficulty	4:09	96	Average	6:07	115	Average	7:01
7	97	Average	6:06	88	Average	6:01	114	Average	6:11
3	97	Average	6:06	88	Average	6:01	90	Average	6:02
11	85	Average	5:09	130	Excellent	6:10	>130	Outstanding	7:01
7	121	Excellent	6:10	105	Average	6:04	120	Above Average	6:09
12	112	Average	6:10	104	Average	6:09	105	Average	6:10
8	82	Below Average	6:02	116	Above Average	6:10	72	Severe Difficulty	5:08
7	<70	Very Severe Difficulty	4:11	118	Above Average	6:10	91	Average	6:03
9	116	Above Average	6:10	118	Above Average	6:10	120	Above Average	7:01
7	105	Average	6:08	118	Above Average	6:10	94	Average	6:04
12	116	Above Average	6:10	93	Average	6:04	105	Average	6:09
9	116	Above Average	6:10	118	Above Average	6:10	120	Above Average	7:01
10	116	Above Average	6:10	106	Average	6:09	114	Average	6:11
8	<70	Very Severe Difficulty	4:09	118	Above Average	6:10	114	Average	6:11
6	105	Average	6:08	118	Above Average	6:10	104	Average	6:08
5	110	Average	6:08	111	Average	6:07	101	Average	6:02
4	103	Average	6:06	79	Severe Difficulty	4:10	102	Average	6:02
9	101	Average	6:08	116	Above Average	6:10	92	Average	6:06
12	112	Average	6:10	116	Above Average	6:10	105	Average	6:10
11	116	Above Average	6:10	118	Above Average	6:10	120	Above Average	7:01
11	116	Above Average	6:10	118	Above Average	6:10	120	Above Average	7:01
7	121	Excellent	6:10	130	Excellent	6:10	125	Excellent	6:10
9	97	Average	6:04	130	Excellent	6:10	100	Average	6:01
7	110	Average	6:08	119	Above Average	6:09	120	Above Average	6:09
14	112	Average	6:10	116	Above Average	6:10	115	Average	7:01
15	<70	Very Severe Difficulty	4:09	116	Above Average	6:10	115	Average	7:01
8	71	Severe Difficulty	5:03	118	Above Average	6:10	120	Above Average	7:01
11	97	Average	6:06	118	Above Average	6:10	96	Average	6:05
10	87	Average	6:02	93	Average	6:04	84	Below Average	6:00
10	116	Above Average	6:10	80	Below Average	5:07	104	Average	6:08
5	116	Above Average	6:10	118	Above Average	6:10	120	Above Average	7:01
6	116	Above Average	6:10	118	Above Average	6:10	99	Average	6:06
12	97	Average	6:06	98	Average	6:07	89	Average	6:02
8	116	Above Average	6:10	106	Average	6:09	114	Average	6:11
8	97	Average	6:04	130	Excellent	6:10	100	Average	6:01

Sound Deletion			Phoneme Awareness		
Std Score	Ability	Age Equiv	Std Score	Ability	Age Equiv
126	Excellent	7:01	104	Average	7:01
106	Average	6:09	97	Average	6:11
111	Average	6:10	95	Average	6:10
104	Average	6:06	90	Average	6:06
90	Average	5:11	84	Below Average	6:02
119	Above Average	6:09	112	Average	6:11
109	Average	6:04	102	Average	6:06
97	Average	6:04	91	Average	6:08
70	Severe Difficulty	5:00	83	Below Average	6:03
99	Average	6:04	96	Average	6:09
120	Above Average	6:11	105	Average	7:00
113	Average	6:10	103	Average	6:11
113	Average	6:10	96	Average	6:09
113	Average	6:10	103	Average	6:11
113	Average	6:10	100	Average	6:11
120	Above Average	6:11	105	Average	7:00
108	Average	6:09	100	Average	6:11
100	Average	5:11	100	Average	6:04
89	Average	5:05	80	Below Average	5:01
93	Average	6:02	91	Average	6:08
101	Average	6:06	95	Average	6:10
108	Average	6:09	100	Average	6:11
129	Excellent	7:01	107	Average	7:01
124	Excellent	6:10	115	Average	6:11
95	Average	5:08	102	Average	6:06
95	Average	5:08	100	Average	6:04
126	Excellent	7:01	104	Average	7:01
126	Excellent	7:01	104	Average	7:01
120	Above Average	6:11	105	Average	7:00
104	Average	6:06	98	Average	6:10
90	Average	5:11	86	Average	6:03
95	Average	6:02	82	Below Average	6:00
120	Above Average	6:11	105	Average	7:00
79	Severe Difficulty	5:05	88	Average	6:04
104	Average	6:06	94	Average	6:08
120	Above Average	6:11	103	Average	6:11
114	Average	6:06	110	Average	6:10

L1 Summary

Child ID	Pre-Age (Y:M)	Post-Age (Y:M)	Gndr	Group	Letter Sound Knowledge					
					Pre-Test			Post-Test		
					Std Score	Ability	Age Equiv	Std Score	Ability	Age Equiv
Q888283008001	6:3	6:7	F	c	116	Above Average	6:10	112	Average	6:10
U888283008032	6:4	6:8	F	c	116	Above Average	6:10	112	Average	6:10
E888283008029	6:1	6:6	F	c	97	Average	6:06	<70	Very Severe Difficulty	4:09
X888283008005	5:11	6:3	F	c	110	Average	6:08	97	Average	6:06
P888283008008	5:8	6:1	F	c	97	Average	6:04	97	Average	6:06
C888283008018	5:6	5:10	F	c	103	Average	6:06	85	Average	5:09
F888283008022	5:6	5:10	F	c	121	Excellent	6:10	121	Excellent	6:10
V888283008023	6:3	6:7	M	c	105	Average	6:08	112	Average	6:10
L888283008035	6:2	6:6	M	c	92	Average	6:04	82	Below Average	6:02
E888283008002	6:0	6:4	M	c	83	Below Average	6:00	<70	Very Severe Difficulty	4:11
Q888283008028	6:0	6:4	M	c	116	Above Average	6:10	116	Above Average	6:10
E888283008031	6:0	6:4	M	c	<70	Very Severe Difficulty	4:09	105	Average	6:08
Z888283008016	5:11	6:3	M	c	110	Average	6:08	116	Above Average	6:10
B888283008027	5:11	6:3	M	c	121	Excellent	6:10	116	Above Average	6:10
L888283008006	5:9	6:2	M	c	<70	Very Severe Difficulty	4:09	116	Above Average	6:10
M888283008026	5:10	6:2	M	c	103	Average	6:06	<70	Very Severe Difficulty	4:09
P888283008010	5:8	6:0	M	c	93	Average	6:02	105	Average	6:08
T888283008012	5:6	5:10	M	c	85	Average	5:09	110	Average	6:08
J803231708113	5:5	5:9	M	c	97	Average	5:09	103	Average	6:06
H888283008004	6:3	6:7	F	i	105	Average	6:08	101	Average	6:08
R888283008019	6:3	6:8	F	i	105	Average	6:08	112	Average	6:10
W888283008014	6:0	6:4	F	i	116	Above Average	6:10	116	Above Average	6:10
F888341408013	5:11	6:3	F	i	121	Excellent	6:10	116	Above Average	6:10
A888283008007	5:7	5:11	F	i	97	Average	6:04	121	Excellent	6:10
H888283008033	5:7	5:11	F	i	103	Average	6:06	97	Average	6:04
X888283008034	5:5	5:9	F	i	>130	Outstanding	6:10	110	Average	6:08
Q888283008030	6:4	6:8	M	i	116	Above Average	6:10	112	Average	6:10
K888358208003	6:2	6:6	M	i	87	Average	6:02	<70	Very Severe Difficulty	4:09
C888283008020	6:0	6:4	M	i	<70	Very Severe Difficulty	4:09	71	Severe Difficulty	5:03
R888283008021	6:1	6:5	M	i	105	Average	6:08	97	Average	6:06
K888283008015	5:11	6:3	M	i	85	Average	5:09	87	Average	6:02
D888283008009	5:10	6:2	M	i	97	Average	6:04	116	Above Average	6:10
J888283008024	5:10	6:2	M	i	121	Excellent	6:10	116	Above Average	6:10
G888283008013	5:9	6:1	M	i	103	Average	6:06	116	Above Average	6:10
Y888283008025	5:9	6:2	M	i	103	Average	6:06	97	Average	6:06
N888283008017	5:8	6:0	M	i	110	Average	6:08	116	Above Average	6:10
U888283008003	5:5	5:9	M	i	109	Average	6:04	97	Average	6:04

YARK Test Results

Sound Isolation						Early Word Recognition					
Pre-Test			Post-Test			Pre-Test			Post-Test		
Std Score	Ability	Age Equiv	Std Score	Ability	Age Equiv	Std Score	Ability	Age Equiv	Std Score	Ability	Age Equiv
118	Above Average	6:10	116	Above Average	6:10	120	Above Average	7:01	115	Average	7:01
118	Above Average	6:10	116	Above Average	6:10	120	Above Average	7:01	115	Average	7:01
106	Average	6:09	96	Average	6:07	96	Average	6:05	115	Average	7:01
92	Average	5:07	88	Average	6:01	108	Average	6:04	114	Average	6:11
97	Average	5:11	88	Average	6:01	94	Average	5:10	90	Average	6:02
111	Average	6:07	130	Excellent	6:10	>130	Outstanding	7:01	>130	Outstanding	7:01
97	Average	5:11	105	Average	6:04	109	Average	6:05	120	Above Average	6:09
106	Average	6:09	104	Average	6:09	94	Average	6:04	105	Average	6:10
98	Average	6:07	116	Above Average	6:10	78	Severe Difficulty	5:08	72	Severe Difficulty	5:08
98	Average	6:07	118	Above Average	6:10	94	Average	6:04	91	Average	6:03
98	Average	6:07	118	Above Average	6:10	110	Average	6:10	120	Above Average	7:01
118	Above Average	6:10	118	Above Average	6:10	96	Average	6:05	94	Average	6:04
105	Average	6:04	93	Average	6:04	110	Average	6:05	105	Average	6:09
119	Above Average	6:09	118	Above Average	6:10	112	Average	6:06	120	Above Average	7:01
111	Average	6:07	106	Average	6:09	125	Excellent	6:10	114	Average	6:11
105	Average	6:04	118	Above Average	6:10	122	Excellent	6:10	114	Average	6:11
111	Average	6:07	118	Above Average	6:10	101	Average	6:02	104	Average	6:08
84	Below Average	5:03	111	Average	6:07	99	Average	6:00	101	Average	6:02
95	Average	5:05	79	Severe Difficulty	4:10	107	Average	5:11	102	Average	6:02
98	Average	6:07	116	Above Average	6:10	94	Average	6:04	92	Average	6:06
118	Above Average	6:10	116	Above Average	6:10	96	Average	6:05	105	Average	6:10
118	Above Average	6:10	118	Above Average	6:10	114	Average	6:11	120	Above Average	7:01
119	Above Average	6:09	118	Above Average	6:10	>130	Outstanding	7:01	120	Above Average	7:01
119	Above Average	6:09	130	Excellent	6:10	108	Average	6:04	125	Excellent	6:10
119	Above Average	6:09	130	Excellent	6:10	99	Average	6:00	100	Average	6:01
103	Average	5:11	119	Above Average	6:09	121	Excellent	6:05	120	Above Average	6:09
118	Above Average	6:10	116	Above Average	6:10	120	Above Average	7:01	115	Average	7:01
118	Above Average	6:10	116	Above Average	6:10	114	Average	6:11	115	Average	7:01
75	Severe Difficulty	5:05	118	Above Average	6:10	102	Average	6:08	120	Above Average	7:01
118	Above Average	6:10	118	Above Average	6:10	90	Average	6:02	96	Average	6:05
111	Average	6:07	93	Average	6:04	85	Average	5:05	84	Below Average	6:00
105	Average	6:04	80	Below Average	5:07	105	Average	6:03	104	Average	6:08
130	Excellent	6:10	118	Above Average	6:10	>130	Outstanding	7:01	120	Above Average	7:01
101	Average	6:01	118	Above Average	6:10	100	Average	6:01	99	Average	6:06
97	Average	5:11	98	Average	6:07	99	Average	6:00	89	Average	6:02
130	Excellent	6:10	106	Average	6:09	125	Excellent	6:10	114	Average	6:11
118	Above Average	6:07	130	Excellent	6:10	102	Average	5:08	100	Average	6:01

Sound Deletion						Phoneme Awareness			
Pre-Test			Post-Test			Pre-Test			
Std Score	Ability	Age Equiv	Std Score	Ability	Age Equiv	Std Score	Ability	Age Equiv	Std Score
99	Average	6:04	126	Excellent	7:01	96	Average	6:09	104
104	Average	6:06	106	Average	6:09	98	Average	6:10	97
104	Average	6:06	111	Average	6:10	96	Average	6:09	95
81	Below Average	5:00	104	Average	6:06	85	Average	5:04	90
95	Average	5:08	90	Average	5:11	92	Average	5:10	84
109	Average	6:04	119	Above Average	6:09	104	Average	6:07	112
89	Average	5:05	109	Average	6:04	89	Average	5:08	102
90	Average	5:11	97	Average	6:04	90	Average	6:06	91
72	Severe Difficulty	5:00	70	Severe Difficulty	5:00	82	Below Average	6:00	83
90	Average	5:11	99	Average	6:04	88	Average	6:04	96
108	Average	6:09	120	Above Average	6:11	96	Average	6:09	105
108	Average	6:09	113	Average	6:10	100	Average	6:11	103
124	Excellent	6:10	113	Average	6:10	108	Average	6:09	96
119	Above Average	6:09	113	Average	6:10	110	Average	6:10	103
114	Average	6:06	113	Average	6:10	106	Average	6:08	100
119	Above Average	6:09	120	Above Average	6:11	106	Average	6:08	105
100	Average	5:11	108	Average	6:09	100	Average	6:04	100
71	Severe Difficulty	4:07	100	Average	5:11	77	Severe Difficulty	4:11	100
<70	Very Severe Difficulty	4:07	89	Average	5:05	83	Below Average	4:11	80
99	Average	6:04	93	Average	6:02	92	Average	6:07	91
85	Average	5:08	101	Average	6:06	90	Average	6:06	95
108	Average	6:09	108	Average	6:09	100	Average	6:11	100
114	Average	6:06	129	Excellent	7:01	108	Average	6:09	107
100	Average	5:11	124	Excellent	6:10	102	Average	6:06	115
71	Severe Difficulty	4:07	95	Average	5:08	94	Average	6:00	102
95	Average	5:05	95	Average	5:08	96	Average	5:08	100
129	Excellent	7:01	126	Excellent	7:01	107	Average	7:01	104
113	Average	6:10	126	Excellent	7:01	103	Average	6:11	104
95	Average	6:02	120	Above Average	6:11	80	Below Average	5:10	105
79	Severe Difficulty	5:05	104	Average	6:06	88	Average	6:04	98
81	Below Average	5:00	90	Average	5:11	94	Average	6:00	86
95	Average	5:08	95	Average	6:02	96	Average	6:02	82
81	Below Average	5:00	120	Above Average	6:11	98	Average	6:03	105
109	Average	6:04	79	Severe Difficulty	5:05	100	Average	6:04	88
105	Average	6:02	104	Average	6:06	96	Average	6:02	94
100	Average	5:11	120	Above Average	6:11	104	Average	6:07	103
101	Average	5:08	114	Average	6:06	104	Average	6:03	110

Post-Test		Extra Tests			
		Split Vowel		Digit Span	
Ability	Age Equiv	Pre-Score	Post-Score	Pre-Score	Post-Score
Average	7:01	5	5	9	11
Average	6:11	2	5	8	8
Average	6:10	3	5	10	14
Average	6:06	2	2	6	7
Below Average	6:02	1	1	5	3
Average	6:11	5	5	10	11
Average	6:06	3	3	8	7
Average	6:08	5	5	9	12
Below Average	6:03	1	1	7	8
Average	6:09	3	2	7	7
Average	7:00	5	4	8	9
Average	6:11	1	4	10	7
Average	6:09	0	3	6	12
Average	6:11	1	5	6	9
Average	6:11	5	4	8	10
Average	7:00	2	5	12	8
Average	6:11	2	5	7	6
Average	6:04	3	1	5	5
Below Average	5:01	0	4	2	4
Average	6:08	1	2	8	9
Average	6:10	1	5	10	12
Average	6:11	1	4	11	11
Average	7:01	5	5	10	11
Average	6:11	3	5	9	7
Average	6:06	1	0	6	9
Average	6:04	1	3	10	7
Average	7:01	5	5	12	14
Average	7:01	3	5	10	15
Average	7:00	5	5	8	8
Average	6:10	1	5	4	11
Average	6:03	0	0	9	10
Below Average	6:00	1	2	7	10
Average	7:00	5	5	5	5
Average	6:04	3	3	8	6
Average	6:08	2	3	7	12
Average	6:11	2	5	6	8
Average	6:10	2	1	5	8
	Mean	2.46	3.57	7.78	8.95
	STD	1.69	1.66	2.26	2.82